APPLICATION OF AN ENSEMBLE-TRAINED SOURCE APPORTIOMENT METHOD TO SPECIATED $PM_{2.5}$ DATA AT THE ST. LOUIS MIDWEST SUPERSITE

A Thesis

Presented to

The Academic Faculty

by

Marissa L. Maier

In Partial Fulfillment

of the Requirements for the Degree

Masters of Science in Engineering in the

School of Georgia Institute of Technology

Georgia Institute of Technology

August 2012

Application of an Ensemble-Trained Source Apportionment Methodology

at the St. Louis Supersite

Approved by:

Dr. Armistead G. Russell, Advisor

School of Civil and Environmental Engineering

Georgia Institute of Technology

Dr. Rodney Weber

School of Earth and Atmospheric Sciences

Georgia Institute of Technology

Dr. Dr. James Mulholland

School of Civil and Environmental Engineering

Georgia Institute of Technology

Dr. Michael H. Bergin

School of Civil and Environmental Engineering & Earth and Atmospheric Sciences

Georgia Institute of Technology

Date Approved: May 18, 2012

Acknowledgements

First and foremost I would like to thank my advisor, Dr. Armistead G. Russell. His wisdom and technical expertise have helped me to grow as a researcher and as an environmental engineer. I would also like to thank Dr. Stefanie Sarnat, Dr. James Schauer, and Dr. Jay Turner for providing sample data and model input files for my analysis. Additionally, I would like to offer a heartfelt thanks to Sivaraman Balachandran and all of the other members of Dr. Russell's research group. You've been a wonderful sounding board for new ideas and have made this entire experience far more enjoyable. Last, but certainly not least I would like to thank my family for their unending support. During the last two years, my husband, Jonathan Walsh, has spent more hours in the airport than most people will in their entire lives so that I could have this wonderful experience.

Table of Contents

List of Tables

List of Figures

List of Symbols

χ^2	Chi-squared
C_{ij}	Concentration of species i on day j
E_i	Emission rate
e_{ij}	Error associated with species i on day j
f_{ik}	Mass fraction of species i in source k
J	Jacobian Matrix
R_i	Chemical reaction term for species i
$\sigma^2_{c_{ji}}$	Uncertainties in the concentration of species i on day j
$\sigma^2_{f_{ik}}$	Uncertainty in the mass fraction of species i in source k
S_{ij}	Sensitivity of species i to parameter j
S_{kj}	Contribution of source k to concentrations on day j
$\bar{S}_j(t_k)$	Ensemble-average impact of source j at time t_k
S	Removal flux
S	Removal flux
u	Velocity

List of Abbreviations

CMAQ	Community Multiscale Air Quality Model
CMAQ-TR	Community Multiscale Air Quality Model Tracer Method
CMB	Chemical Mass Balance
CMB-MM	Chemical Mass Balance Molecular Markers
DDM	Direct Decoupled Method
EBSPs	Ensemble-Based Source Profiles
EC	Elemental Carbon
lpm	Liter per Minute
MBSPs	Measurement-Based Source Profiles
mm	Millimeters
OC	Organic Carbon
PM	Particular Matter
PMF	Positive Matrix Factorization
PMF-MM	Positive Matrix Factorization Molecular Markers
S/N	Signal-to-noise
SA	Source Apportionment
SOC	Secondary Organic Carbon
STL-SS	St. Louis Supersite
STL-SS	St. Louis Supersite
XRF	X-ray Fluorescence
$\mu g/m^3$	Micrograms per Cubic Meter

Summary

Five different receptor and chemical transport models were used to quantify the sources of $PM_{2.5}$ impacting the St. Louis Supersite (STL-SS) between June 2001 and May 2003. Since each source apportionment (SA) technique has its own limitations, this work compared the results of five different SA approaches to better understand the biases and limitations of each. Additionally, this work incorporated the source impacts predicted by these five models into an ensemble-trained SA methodology developed by Lee et al. 2009. The ensemble method offered several improvements over the five individual SA techniques. Primarily, the ensemble method was able to calculate source impacts on days when individual models either did not converge to a solution or did not have adequate input data to develop source impact estimates. Additionally, the ensemble method resulted in fewer days when major emissions sources (e.g., secondary organic carbon and diesel vehicles) were estimated to have either zero impact or a negative impact on $PM_{2.5}$ concentrations at the STL-SS. When compared with a traditional chemical mass balance (CMB) approach using measurement-based source profiles (MBSPs), the ensemble method was associated with better fit statistics, including reduced chi-squared values and improved $PM_{2.5}$ mass reconstruction.

A comparison of the different modeling techniques also revealed some of the subjectivities associated with applying specific SA models to the STL-SS dataset. For instance, positive matrix factorization (PMF) results were sensitive to both the fitting species and number of factors selected for the analysis. Moreover, source impacts predicted in CMB were sensitive to the selection of representative metals processing profiles. This was associated with a considerable amount of uncertainty at the STL-SS since the metals processing point sources affecting the monitor were not well-characterized. These issues highlighted the value of using several different SA techniques at a given receptor site, either by comparing source impacts predicted by different models or by utilizing an ensemble-trained SA technique.

Chapter 1: Introduction

Numerous epidemiological studies have demonstrated an association between ambient air pollution and cardiorespiratory disease. A detailed understanding of the relationship between emission sources and pollutant concentrations may help to provide further insight into the impact of specific emissions sources on human health. However, there are several issues that limit the use of current source apportionment (SA) methods in epidemiological analyses. First, current receptor modeling techniques cannot identify and evaluate all emissions sources. For example, Baek et al. (2005) found that typical receptor model applications only include/identify sources representing 60-80% of inventoried primary emissions (Baek, Park et al. 2005). Secondly, certain SA methods appear to introduce excessive day-to-day variability in source impacts. This may result in little to no predicted impact from a major emissions source, such as gasoline vehicles, on a given day followed by high predicted impacts the next day. Conversely, other SA methods appear to under-predict day-to-day variability in source impacts, limiting their applicability to health studies that associate acute responses to increases in pollutant levels.

In addition to these limitations, it is difficult to evaluate SA model performance and calculate uncertainties because source contributions at a specific receptor location cannot be directly measured. One method of dealing with the limitations of any single SA approach is to apply a number of different SA methods to the same dataset. Different SA models may be compared to one another to better understand the biases and uncertainties associated with specific techniques. While consistency between methods does not necessarily assure accurate results, a detailed evaluation of the circumstances in which the model results converge or diverge can still be informative. Alternatively, Lee et al. (2009) developed a method of combining results from different SA models to find optimized source profiles and contributions.

The objective of this work was to evaluate the emissions sources that contribute to ambient particulate matter less than 2.5 microns in aerodynamic diameter ($PM_{2.5}$) at the St. Louis Supersite (STL-SS) in East St. Louis, Missouri. In order to accomplish this, a number of different SA techniques were applied to: (1) identify and characterize the emissions sources affecting the STL-SS; (2) quantify the contributions of these emissions sources to total $PM_{2.5}$ concentrations at the STL-SS; and (3) assess the uncertainties associated with the predicted source contributions.

Chapter 2: Background

Source Apportionment Models

Source apportionment models quantify the contributions of various pollutant sources to pollutant concentrations at a specific location. Most widely used source apportionment techniques (e.g., Chemical Mass Balance, Positive Matrix Factorization) fall under the category of receptor models, which rely on monitoring data to evaluate pollutant sources affecting a receptor location. However, photochemical or dispersion models can also be used to assess the impact of various emissions sources at a specific location. These models utilize emissions data, meteorological data, and mathematical representations of chemical transformation mechanisms to provide spatially and temporally varying pollutant concentrations fields (USEPA 2010).

Chemical Mass Balance

The Chemical Mass Balance (CMB) method is a receptor model that utilizes the chemical and physical characteristics of pollutants measured at an emissions source and a receptor site to: (1) identify the presence of various emissions sources at the receptor location; and (2) quantify daily contributions of each emission source to measured receptor concentrations. CMB assumes that the ambient concentration of a chemical species may be expressed as the sum of the product of source profiles and source contributions:

Equation 1

$$C_{ij} = \sum_{k=1}^{M} f_{ik} \cdot S_{kj} + e_{ij}$$

In Equation 1, above, C_{ij} is the observed concentration of species i on day j, f_{ik} is the mass fraction of species i in source k, S_{kj} is the contribution of source k to ambient concentrations on day j, M is the number of sources, and e_{ij} is the error term associated with species i on day j.

User inputs consist of ambient concentration data (C_{ij}), measurement-based source profiles (f_{ik}), and associated uncertainties. The CMB model then calculates the contribution of each identified emission source to measured particulate concentrations on a given day (S_{kj}) (Coulter 2004). CMB employs a weighted effective variance technique (Watson, Cooper et al. 1984) to solve for the source contributions that minimize the weighted sum of the squares of the differences between measured and calculated species concentrations:

Equation 2

$$\chi_j^2 = \sum_{i=1}^{M} \frac{\left(C_{ij} - \sum_{k=1}^{N} f_{ki} S_{kj}\right)^2}{\sigma_{c_{ji}}^2 + \sum_{k=1}^{N} \sigma_{f_{ik}}^2 S_{kj}^2}$$

In Equation 2, the difference between the measured and predicted concentrations is weighted by the uncertainties associated with the ambient concentration data ($\sigma_{c_{ji}}^2$) and the species profile ($\sigma_{f_{ik}}^2$). This technique gives greater influence to chemical species with small measurement and source impact

uncertainties than to species with large uncertainties. Additionally, CMB uses an iterative approach to estimate the uncertainties associated with daily source contributions estimated by the model (Coulter 2004):

Equation 3

$$\sigma_{s_{jk}} = \sum_{j=1}^{N} \left[\frac{f_{ki}^2}{\sigma_{c_{ji}}^2 + \sum_{k=1}^{N} \sigma_{f_{ik}}^2 s_k^2} \right]^{-1/2}$$

The uncertainties associated with average source contributions were estimated using the root mean square average of daily uncertainties in the source contributions.

The CMB model formulation is based on several assumptions that limit its accuracy and applicability. First, Equation 1 assumes that the species do not undergo any sort of chemical transformations between the emissions source and the receptor location. Thus, CMB is unable to explicitly account for secondary pollutant formation and the chemical loss of individual compounds. Additionally, the model assumes that the number of source profiles is less than or equal to the number of species, the sources are linearly independent, and the measurement uncertainties are random, uncorrelated, and normally distributed (Coulter 2004). Additionally, model accuracy is highly dependent upon the user's ability to account for and characterize all emissions sources affecting the receptor location. Further, the model assumes that these source emissions have a constant chemical composition (Coulter 2004). The last two assumptions may introduce significant errors, since source profiles may change with time. Further, there can be significant variations in source composition by region, day-of-week, season, and/or ambient conditions (e.g., changes in gasoline composition and engine operating temperature) (Marmur, Mulholland et al. 2007). Further, if the source profiles are linearly dependent or close to linearly dependent, errors in the source profiles or measurements can lead to large errors in source impact estimates.

CMB is typically used to reconstruct $PM_{2.5}$ mass, based on speciated $PM_{2.5}$ data, consisting of ions (e.g., sulfate, nitrate, ammonium), organic carbon, elemental carbon, and trace elements (e.g., calcium, lead, zinc). However, CMB may also be used to reconstruct organic carbon mass, using measurement data for particle-bound organic compounds, which serve as molecular markers. This approach has been applied by Bae and others in St. Louis and elsewhere (Bae, Schauer et al. 2006).

Factor Analytic Approaches

A second class of receptor models use factor analytical approaches to decompose an (j x i) matrix of measurement data into factors profiles and factor contributions. Commonly used factor analytical methods include Principal Component Analysis, Unmix (USEPA 2010), and Positive Matrix Factorization (PMF) (USEPA 2008). PMF was the factor analytical tool used in this work. Like CMB, PMF assumes that the concentration of species i can be expressed as a linear combination of source profiles and source contributions (Equation 1). PMF solves for the source profiles and source contributions that minimize

the object function Q over a time series of ambient measurement data, such that no sample has a negative source contribution[1]:

Equation 4

$$Q = \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{C_{ij} - \sum_{k=1}^{P} f_{ik} S_{kj}}{\sigma_{C_{ij}}}$$

As shown in Equation 4, PMF allows each data point to be individually weighted by its associated uncertainty ($\sigma_{C_{ij}}$), so that certain values, such as below detection limit measurements, may be adjusted to have less influence on the solution. Additionally, the user can adjust the uncertainties associated with the ambient concentration data input into the model by categorizing each species as either "strong", "weak", or "bad". These assignments are typically made from the signal-to-noise ratio (S/N) calculated for each species. The PMF Users Guide recommends that species with S/N less than 0.2 be categorized as bad species and excluded from model calculations. Species with S/N between 0.2 and 2 are typically categorized as weak species, and the user input measurement uncertainties associated with these species are tripled. Species with S/N greater than 2 are typically categorized as strong species. The measurement uncertainties associated with strong species are not adjusted (USEPA 2008).

PMF assesses the stability and uncertainty of its model output using a bootstrapping technique. PMF performs bootstrapping by randomly selecting non-overlapping blocks of samples and creating a new input data file from the selected samples, with the same dimensions as the original dataset. PMF is then run on the new data sets and each bootstrapping factor is mapped to a base run factor by comparing factor contributions (USEPA 2008). Uncertainties in the output factor profiles can be estimated as the standard deviation of 100 bootstrapping runs. The uncertainty in the factor contributions of species i is calculated as the product of the factor contributions (s_{kj}) times the uncertainty in the factor profile:

Equation 5

$$\sigma_{ij}^2 = \sum_k \sigma_{f_{ik}}^2 S_{kj}^2$$

One distinct advantage of factor analytical approaches, such as PMF, is that the user is not required to input measurement-based source profiles (MBSPs). This allows PMF to be used in situations where the emissions sources affecting the receptor location are not well-characterized. However, factor analytical approaches do rely upon the user to select an appropriate number of factors (profiles) and link these factors to specific emissions sources. Thus, the application of factor analytic models may be somewhat subjective. Further, since PMF solves for source profiles over an entire dataset (rather than on a day-by-day basis), this approach does not work well for small datasets. Thus, PMF is not typically used on datasets with less than 100 days of measurement data (USEPA 2008).

As with CMB, PMF is typically run on speciated $PM_{2.5}$ datasets using major ions, elements and EC/OC concentrations. When available, EC and OC fractions (e.g., OC1-4, OP, and EC1-3) can also be used. Further, PMF may also be used to apportion organic carbon mass from molecular marker datasets (PMF-

[1] In practice, sources are allowed to have small negative contributions to $PM_{2.5}$ mass.

MM). PMF-MM has been used by Jaeckels et al. and others on molecular marker data sets collected in St. Louis and in other locations (Jaeckels, Bae et al. 2007).

Emissions-Based Models

In addition to receptor-based approaches, emissions-based chemical transport models (CTMs) have been used for SA. CTMs use species conservation equations to estimate pollutant concentrations over a series of three dimensional grids, based on emission estimates and meteorological data:

Equation 6

$$\frac{\partial c_i}{\partial t} = -\nabla(\mathrm{uc}_i) + \nabla \mathrm{K} \nabla \mathrm{C}_i + \mathrm{R}_i(c_1, c_1, \ldots, c_n) + E_i - S$$

In the above equation, c_i is the concentration of species i, u is the velocity vector, R_i is the chemical reaction term, E_i is the emission estimate for species i, and S is the removal flux or sink term.

In order to use CTMs for SA, it is necessary to determine how emissions from various sources impact concentrations at specific receptor locations. Sensitivity analyses, such as the decoupled direct method (DDM-3D), are commonly used to link receptor concentrations with emissions rates. More broadly, DDM-3D investigates the response of atmospheric concentrations, $c_i(x,t)$, to perturbations in model parameters or inputs, such as emission rates (Baek 2009). DDM-3D computes the first-order semi-normalized sensitivity of c_i to perturbations in emissions by solving the following equation:

Equation 7

$$\frac{\partial S_{ij}}{\partial t} = -\nabla\left(uS_{ij}\right) + \nabla\left(K \nabla S_{ij}\right) + JS_{k,j} + E_i'$$

In Equation 7, S_{ij} is the sensitivity of species i to parameter j and E_i is the unperturbed emission rate. J is the Jacobian matrix, which represents the chemical interaction between species (Baek 2009). Unlike, CMB and PMF, the emissions-based source apportionment model used in this work does not assess uncertainties in predicted source impacts. Thus, source impact uncertainties were estimated using an approach developed by Lee et al. 2009, where CTM-based source impacts were compared to impacts predicted using another SA technique.

Unlike other SA methodologies, CTMs can evaluate how individual sources impact secondary pollutant formation (e.g., sulfate, nitrate, and secondary organic aerosol production). CTM-based SA methods also provide the distinct advantage of having greater spatial coverage than traditional receptor-based SA models (Lee, Balachandran et al. 2009). Despite these advantages, there are still a number of limitations associated with the use of CTMs for SA. Photochemical models are computationally intensive, time-consuming to implement, and have extensive input data requirements (e.g., meteorological and emissions data). Further, emissions data are generally available as spatial and temporal averages. Thus, space and time resolved emission rates must be estimated from average emissions inventory data based on other variables like local land information. Since day-to-day variability in emissions and transport of the emissions to the receptor is not well characterized, CTM-based SA results often show very little variation in modeled source impacts compared to receptor modeling techniques (Stolzel, Laden et al.

2005; Hogrefe, Porter et al. 2006; Marmur, Park et al. 2006; Lee, Balachandran et al. 2009). This limitation of the use of CTMs for SA is substantial when SA results are to be used in time –series health studies.

Ensemble-Trained Source Apportionment

The ensemble-trained SA technique developed by Lee et al. 2009 blends receptor and chemical transport model outputs to develop ensemble-based source profiles (EBSPs). These EBSPs can then be used in CMB to refine estimated source impacts at a selected receptor location.

The technique developed by Lee et al. ensembles source contributions from a number of different SA models using a weighted average:

Equation 8

$$\bar{S}_j(t_k) = \frac{\sum_{l=1}^{L} w_{jl}(t_k) \cdot S_{lj}(t_k)}{\sum_{l=1}^{L} w_{jl}(t_k)}$$

Where $\bar{S}_j(t_k)$ is the ensemble-average impact of source j (in µg/m^3) at time t_k, $S_{lj}(t_k)$ is the impact estimated from method l (e.g., CMAQ-TR, PMF, PMF-MM), and $w_{jl}(t_k)$ is the weight assigned to the calculated source impact. Lee et al. used the inverse of the uncertainties associated with individual source impacts ($\sigma_{S_{ij}}$) as weighting factors. However, if $w_{jl}$is set equal to one, $\bar{S}_j(t_k)$ becomes the arithmetic mean of the daily source impact estimates.

Lee et al. averaged source impact estimates from five different models during a one month period in the winter and a one month period in the summer to account for seasonal variability. The ensemble-averaged source impacts were then used to calculate optimized summer and winter ensemble-based source profiles, using a CMB approach that minimized chi-squared (Equation 2). The optimized winter and summer source profiles were then used to develop updated source impacts over a longer time series of data.

In theory, the ensemble-trained SA technique should blunt the problems associated with any one specific technique. Lee et al. demonstrated that the ensemble technique reduced the frequency of zero-impacts days for major emissions sources (e.g., gasoline vehicles) when applied to a 10 year time series of speciated $PM_{2.5}$ data in Atlanta, compared to a traditional CMB application. Performance measures, such as chi-squared and ratios of predicted-to-observed $PM_{2.5}$ ratios also indicate improved performance using the EBSPs in CMB compared to the MBSPs for the same Atlanta receptor (Lee, Balachandran et al. 2009). Further, the ensemble method calculates EBSPs from average source impacts estimated using a number of different SA techniques; therefore, this method should also reduce the uncertainty associated with identifying a representative set of MBSPs. One major limitation of ensemble-trained SA is the time associated with applying multiple SA techniques to a single receptor location. Further, since the ensemble method ultimately utilizes CMB to calculate source impacts using optimized source profiles, this approach is still subject to some of the same limitations as CMB.

Chapter 3: Methods

In this work, five different SA techniques were applied to assess source contributions to measured $PM_{2.5}$ concentrations at the STL-SS between June 2001 and May 2003. The source contributions predicted by the five different SA techniques were then compared with one another to better understand the advantages and disadvantages associated with each technique. The uncertainties associated with each source apportionment technique were also evaluated, and the results from the five SA methods were incorporated into an ensemble-trained SA model.

Monitoring data

The USEPA-funded St. Louis Midwest Supersite was located in a low density, urban residential/light commercial area in East St. Louis, approximately 3 km east of the St. Louis' central business district (Figure 1). This site is impacted by numerous industrial sources, several of which are shown in Figure 1. Further, a review of EPA's National Emissions Inventory for 2002 revealed a lead smelter in Jefferson County, a zinc smelter and copper processing facility in St. Clair County, and a steel processing facility and copper smelter in Madison County (USEPA 2010).

Daily, filter-based fine particulate matter samples were collected at the STL-SS between June 2001 and May 2003. These samples were collected from midnight-to-midnight Central Standard Time, and analyzed for fine particle mass, OC, EC, sulfate, nitrate, ammonium, trace elements, and organic tracers (Bae, Schauer et al. 2006). Fine particulate mass and elemental composition were measured using a sampling system consisting of a Harvard Impactor, operating at a flow rate of 10 liters per minute (lpm), and a 37 mm Teflon filter. After $PM_{2.5}$ mass concentrations were determined by gravimetric analysis, the filters were analyzed by energy-dispersive X-ray fluorescence (XRF) for 40 trace elements. Major ions (sulfate, nitrate, and ammonium) were measured using a Harvard-EPA Annular Denuder System (HEADS), consisting of a glass inlet, an impactor, two glass annular denuders, a 47 mm Teflon filter, and a 47 mm Nylon filter. The HEADS was run at a flow rate of 10 lpm (Lee, Hopke et al. 2006).

Organic and elemental carbon were measured using the University of Wisconsin low-volume sampling train, consisting of two parallel organic OC/EC sampling channels operating at a flow rate of 12 lpm. The first channel consisted of an organics denuder followed by a 27 mm prebaked quartz filter. EC and OC concentrations were measured from the quartz filters using two separate methods: the National Health and Safety/ Thermal Optical Transmittance method and the IMPROVE/ Thermal Optical Reflectance method (Lee, Hopke et al. 2006).

OC, EC, and particle-bound organic tracers were also measured on 90 mm quartz filters, using a 92-lpm, medium volume sampler equipped with a $PM_{2.5}$ cyclone. 1-in-6 day samples were analyzed by solvent extraction gas chromatograph mass spectroscopy (GCMS), while the remaining sample days were analyzed using thermal desorption GCMS (Sheesley, Schauer et al. 2007). Target analytes for the 1-in-6 day samples included n-alkanes, cycloalkanes, alkanoic acids, resin acids, aromatic diacids, alkanedioic acids, steranes, hopanes, PAHs, oxy-PAHs, phthalates, levoglucosan, and cholesterol.

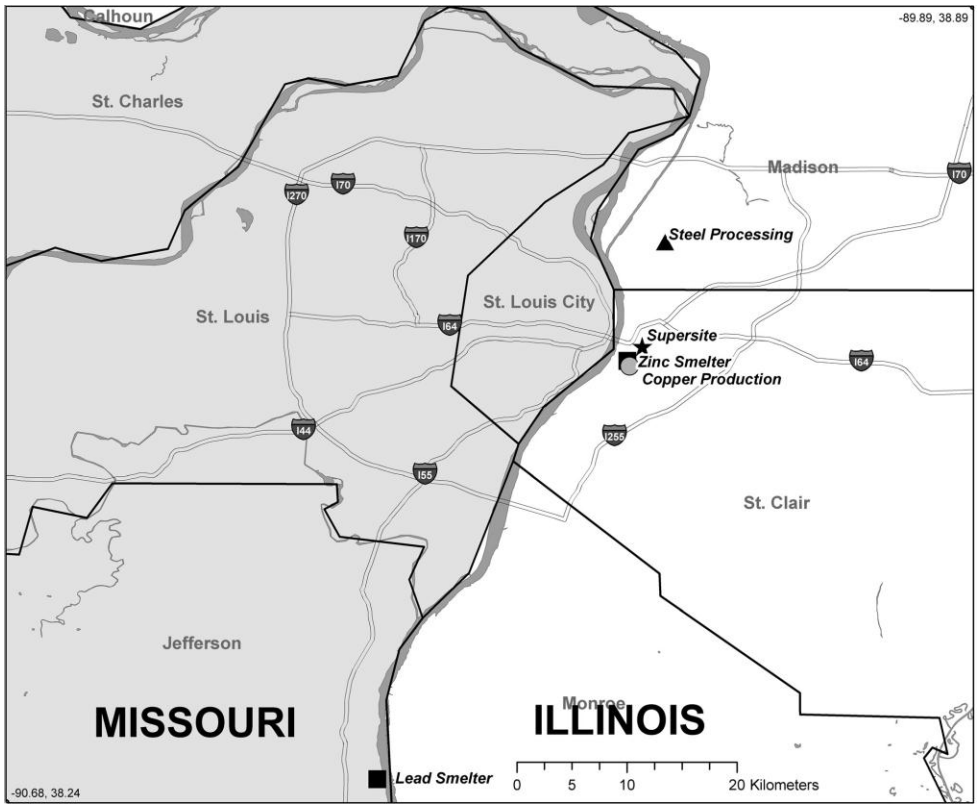


Figure 1: Location of the St. Louis Midwest Supersite

Source Apportionment Results

The five techniques used for source apportionment at the STL-SS were positive matrix factorization (PMF), positive matrix factorization using molecular markers (PMF-MM), chemical mass balance (CMB), chemical mass balance using molecular markers (CMB-MM), and the Community Multiscale Air Quality Model tracer method (CMAQ-TR). Source contributions predicted by PMF, PMF-MM, CMB-MM, and CMAQ-TR were combined to find winter and summer EBSPs. The EBSPs were then used to develop revised source impact estimates based on the ensemble-trained SA technique developed by Lee et al. 2009.

Positive Matrix Factorization

Several researchers have used PMF to estimate source impacts at the STL-SS. Lee et al. (2006) ran PMF on a speciated $PM_{2.5}$ dataset that included IMPROVE OC and EC fractions, ions (sulfate, nitrate, ammonium), and XRF metals. Lee et al. determined that ten PMF factors produced a robust solution, which corresponded well with *a priori* knowledge of the $PM_{2.5}$ sources in East St. Louis. The factors and average source contributions were: secondary sulfate (32.6%), carbon-rich sulfate (19.6%), gasoline exhaust (16.4%), secondary nitrate (15.3%), steel processing (6.8%), airborne soil (4.2%), diesel emissions/ railroad traffic (2.1%), zinc smelting (1.3%), lead smelting (1.3%), and copper production (0.5%) (Lee, Hopke et al. 2006). Garlock 2006 used PMF to refine Lee et al.'s predicted source impacts at the STL-SS. Garlock's ten factor PMF solution corresponded to the same emissions sources identifed by Lee et al., with similar source contributions (Garlock 2006).

The PMF analysis conducted in this work utilized a similar raw dataset as Lee et al. and Garlock. This dataset included the IMPROVE OC/EC fractions[2] in an effort to resolve separate gasoline and diesel vehicle impacts. The dataset also included sulfate, nitrate, and ammonium ions and 11 of the 21 XRF species considered in the previous source apportionment studies (Al, As, Ca, Cu, Fe, K, Mn, Pb, Se, Si, and Zn). These XRF species were selected in part from a review of method detection limits, analytical uncertainties, and calculated S/N. A sensitivity analysis was also preformed to evaluate the impact of adding or removing fitting species from the PMF input file. This analysis showed that the PMF results for the STL-SS dataset were sensitive to the choice of XRF fitting species. Primarily, when all 21 of the XRF species were included in the PMF input files, PMF was unable to resolve separate gasoline and diesel vehicle factors.

As discussed above, PMF requires concentrations and associated measurement uncertainties for each fitting species used in the analysis. The raw measurement data was processed for input into PMF using the procedures outlined in Reff et al 2007. In general, if a species was measured at a concentration above the detection limit, the concentration was set equal to the measured value and the uncertainty was set equal to the analytical uncertainty plus one third of the detection limit. If a measurement fell below the detection limit, the concentration was set equal to one half the detection limit and the uncertainty was set equal to five sixths the detection limit (Reff, Eberly et al. 2007). Days with missing ion, XRF, organic/elemental carbon, or $PM_{2.5}$ mass measurements were excluded from the dataset (Garlock 2006; Reff, Eberly et al. 2007).

Method detection limits and analytical uncertainties for the STL-SS dataset were obtained from a number of sources. For the IMPROVE OC/EC fractions and the XRF data, detection limits and analytical uncertainties were provided with the Microsoft Excel files containing the measurement data (Turner 2011). Method detection limits for the ionic species were found in the St. Louis – Midwest Fine Particulate Matter Supersite Quality Assurance Final Report. Lastly, in the absence of STL-SS specific data, method detection limits and analytical uncertainties were assumed to be equivalent to those used in the Southeastern Aerosol Research and Characterization (SEARCH) study (Atmopsheric Research & Analysis 2003; Turner 2007). SEARCH method detection limits and analytical uncertainties were applied to the $PM_{2.5}$ mass and ion data.

Calculations were performed using EPA PMF v3.0.2.2. Within PMF, EC3 and As were considered “weak” species with S/N less than two. The uncertainties associated with these species were multiplied by a factor of three to lower their influence on the final PMF solution. 655 days of measurement data were used in the analysis.

Based on the results of the Lee et al. and Garlock studies, this work evaluated both a 10 factor and an 11 factor solution. The 10 factor solution was assigned to the following emissions sources: secondary sulfate, secondary nitrate, motor vehicles, resuspended soil, resuspended soil/road dust, biomass burning, steel processing, industrial copper, industrial lead, and industrial zinc. The most significant difference between the 10 and 11 factor solutions was the splitting of the motor vehicle source into

[2] In order to ensure mass closure, the OP fraction was subtracted from total EC.

separate gasoline and diesel vehicle factors. The gasoline factor from the 11 factor PMF solution contained higher percentages of the lighter OC and EC fractions (OC1 and EC1) than the diesel factor. The gasoline factor also had a lower EC/OC ratio (0.49) than the diesel factor. Conversely, the diesel factor from the 11 factor PMF solution contained more of the heavier OC and EC fractions (OP and EC2) by mass percent, and had a higher EC/OC ratio than the gasoline factor (0.60) (Figure 2).

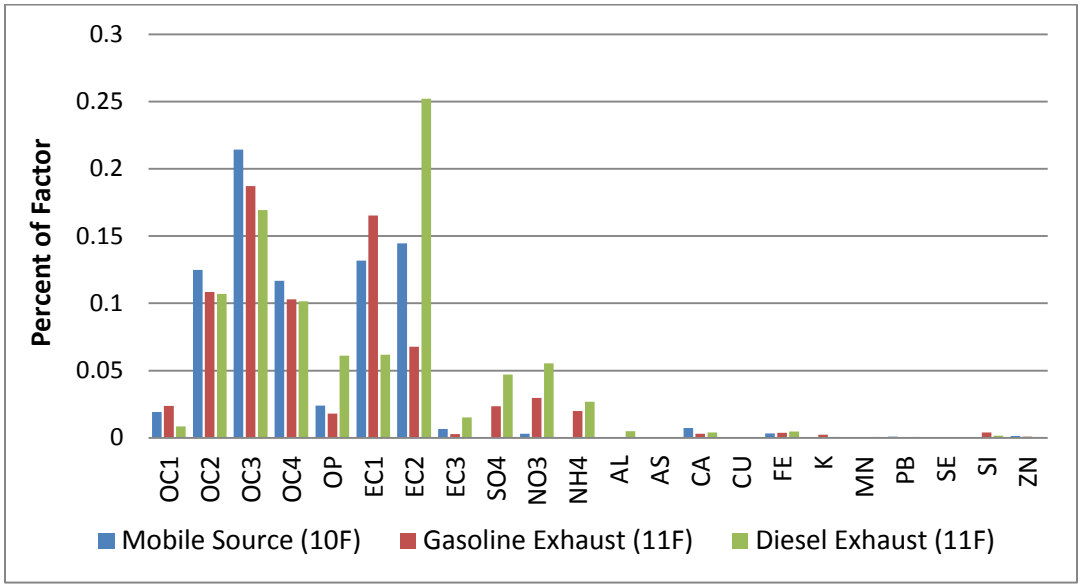


Figure 2: Comparison of Mobile Source Factors

Table 1: Comparison of Model Performance Metrics for the 10 Factor and 11 Factor PMF Solutions

	10 Factor Solution	11 Factor Solution
Modeled/Measured $PM_{2.5}$ Mass	99.0 +/- 10.9%	99.9 +/- 11.3%
Q(Theoretical)	7205	6550
Q(Robust)	16152.9	12201

In order to evaluate model performance, theoretical Q values were calculated for the 10 factor and 11 factor solutions and compared to the robust Q values provided in the model output. The theoretical Q value depends upon the number of data points (n), "good" species (m_g), "weak" species (m_w), and factors (p) (USEPA 2008):

Equation 9

$$Q(theoretical) = n(m_g + m_w/3 - p)$$

As show in Table 1, the theoretical and robust Q values are closer for the 11 factor solution than they are for the 10 factor solution. Further, the 11 factor solution does a slightly better job reconstructing measured $PM_{2.5}$ mass. However, the standard deviation of this ratio is larger for the 11 factor solution than it is for the 10 factor solution. The standard deviation reflects the model's tendency to both under predict and over predict measured $PM_{2.5}$ mass, even if the average value is close to 100%. When both the average and standard deviation of the modeled-to-measured $PM_{2.5}$ ratio are considered, it is difficult

to determine which solution better fits the measurement data. However, based on the Q statistics and the desire to resolve separate gasoline and diesel vehicle factors, the 11 factor solution was incorporated into the ensemble calculations.

Bootstrapping was used to estimate the stability and uncertainty of the 11 factor solution by randomly selecting non-overlapping blocks of samples and creating new input data files from the selected samples, with the same dimensions as the original dataset. PMF was then run on the new data sets and each bootstrapping factor was mapped to a base run factor by comparing factor contributions (USEPA 2008). Factor contributions were determined from the average of 100 bootstrapping runs, whereas uncertainties were determined from the standard deviation of 100 bootstrapping runs. Total source contributions and uncertainties are shown in Figure 3, whereas individual factor compositions are provided in Appendix A.

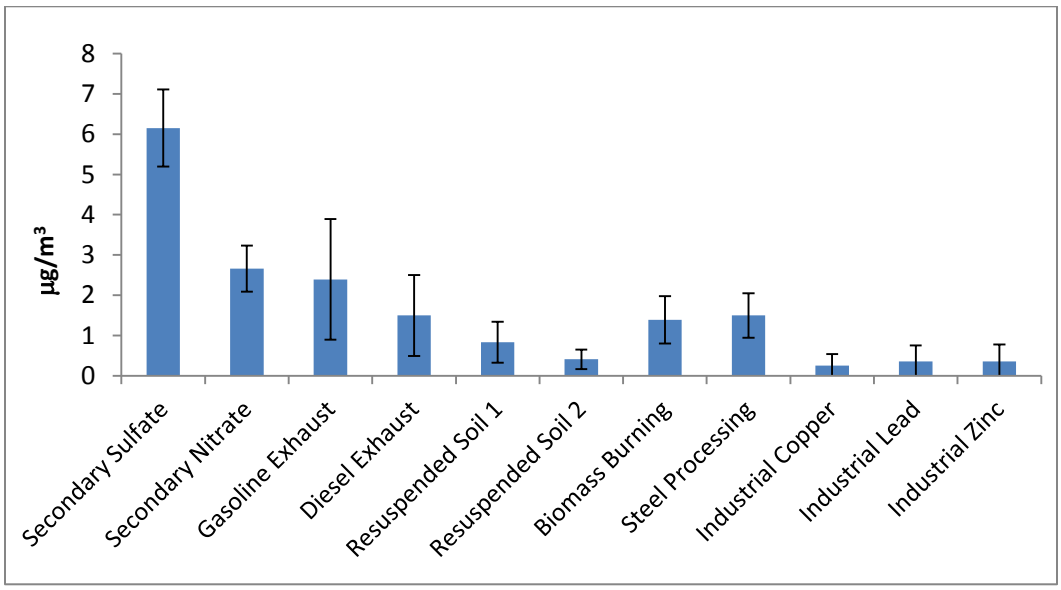


Figure 3: 11 Factor PMF Solution

Positive Matrix Factorization - Molecular Markers

Jaeckels et al. used PMF on an organic molecular marker dataset (PMF-MM) to estimate source contributions to OC concentrations at the STL-SS between May 2001 and June 2003. Calculations were performed using the EPA PMF 1.1 software. Jaeckels et al. used the 1-in-6 day organic molecular marker, OC/EC, silicon, and aluminum measurements collected at the STL-SS as input data for PMF. 112 days with valid measurement data were used in the analysis. The researchers determined that the eight factor PMF solution was the most useful in assessing source contributions to OC at the STL-SS. Based on seasonal trends, day-of-week trends, and the chemical composition of each factor, the following emission source categories were identified: secondary organic aerosol (SOA), wood combustion, winter combustion 1, winter combustion 2, mobile factor, point source 1, point source 2, and resuspended soil. The SOA factor was characterized by alkanedioic acids, phthalates, aromatic diacids, and alkanoic acids, whereas the biomass combustion factor was characterized by levoglucosan, resin acids, and steranes. Both winter combustion factors contributed more to OC mass in the wintertime than in the summertime. The first winter combustion factor was high in PAHs and oxygenated PAHs, while the

second factor was high in cholesterol and PAHs. The mobile source factor was identified by the presence of steranes, hopanes, and EC, while silicon, aluminum, and cholesterol were used to identify the resuspended soil factor. The remaining two PMF factors were characterized by episodic contributions to OC and were classified as point source factors. The first point source factor was dominated by PAHs, oxygenated PAHs, and resin acids, while the second point source is dominated by alkanes, aromatic diacids, and alkanoic acids (Jaeckels, Bae et al. 2007).

In order to obtain daily source contributions, PMF-MM was rerun using PMF-MM input files provided by Turner (Turner 2011). These input files consisted of ambient concentration data and associated measurement uncertainties for 107 particle-bound organic compounds, OC/EC, silicon, and aluminum. Eleven days of measurement data were added to the original data set, including four days in July 2001. Measurement uncertainties for these eleven days were estimated by regressing the initial 112 days of ambient concentration data by their associated measurement uncertainties. The following species had S/N less than two, and were treated as "weak" species in PMF by tripling provided measurement uncertainties: pentadecylcyclohexane, heptadecylcyclohexane, octadecylcyclohexane, nonadecylcyclohexane, pimaric acid, benz(a)anthracene, indeno(c,d)pyrene, and stigmasterol. Based on the description provided in Jaeckels et al., 9-hexadecenoiacid was assigned as a "bad" species and excluded from the PMF analysis.

The resulting eight factor PMF solution was similar to the eight-factor PMF solution reported by Jaeckels et al. One significant different between this work and the results presented in Jaeckels et al. was the identification of two wood burning factors, rather than two winter combustion factors. The first wood burning factor was relatively high in levoglucosan, whereas the second wood burning factor was high in resin acids (see Appendix B for factor profiles). An attempt was made to investigate a 7 factor solution; however, PMF could not find a convergent solution for fewer than 8 factors. Contributions and uncertainties were assessed using the average and standard deviation of 100 bootstrapping runs. Average contributions to OC mass and associated uncertainties are shown in Figure 4, below. On average, PMF-MM was able to reconstruct $98.4 \pm 40.6\%$ of the measured OC mass. Source contributions to $PM_{2.5}$ were estimated based on the source-specific OC to $PM_{2.5}$ ratios.

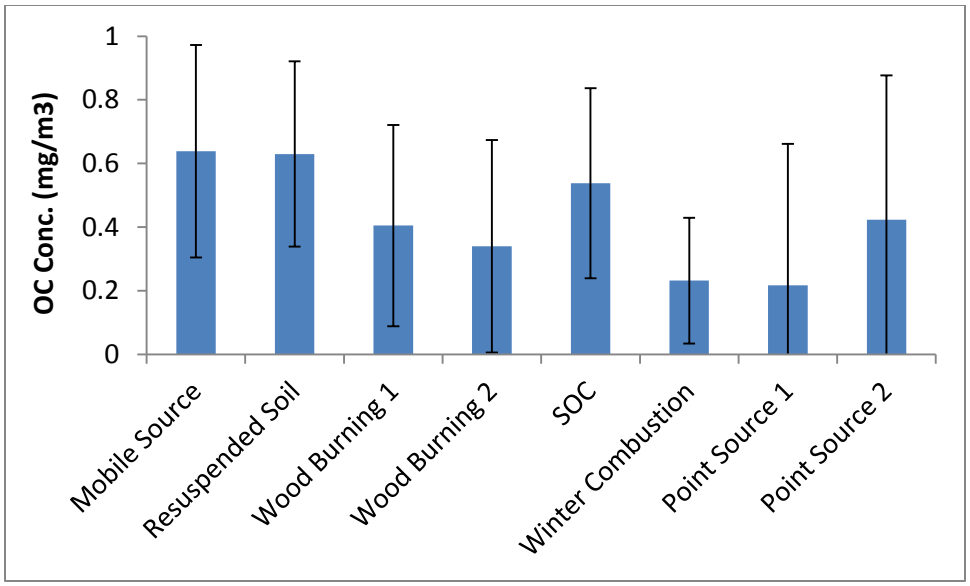


Figure 4: Average OC Contributions and Uncertainties Calculated Using PMF-MM

Chemical Mass Balance

This work also applied CMB to assess source contributions to total $PM_{2.5}$ at the STL-SS. A literature review indicated that CMB has either not yet been applied to the STL-SS dataset or the results of such an analysis have not been published. Inputs to CMB included concentration and uncertainty files, which were developed using the same methodology as the PMF data inputs. The only difference between the concentration and uncertainty data used in CMB and PMF was the treatment of missing ion measurements. Missing ion measurements were replaced with the geometric mean of the measured values and the associated uncertainties were set equal to four times the geometric mean. This data substitution procedure only applied to missing nitrate and ammonium measurements, since days with missing sulfate measures also had other missing measurements and were excluded from the dataset (Garlock 2006; Reff, Eberly et al. 2007). This procedure led to the inclusion of 686 days of measurement data in the CMB analysis.

In addition to ambient concentration data, CMB requires user input source profiles that represent the important emissions sources affecting the receptor location. Source profiles for gasoline vehicles, diesel vehicles, biomass burning, dust, ammonium sulfate, ammonium bisulfate, and ammonium nitrate were included in the model (Marmur, Unal et al. 2005). Since the PMF results indicate that the STL-SS monitor is also impacted by industrial metals (e.g., zinc, copper, iron, manganese, lead), it was necessary to develop representative metals processing profiles for use in CMB. Metals profiles were developed from the MBSPs provided in EPA's Speciate database (USEPA 2011). Average copper and steel processing profiles were selected from the Speciate database. These profiles were composited from measurements of different emissions sources associated with primarily copper smelting and steel processing operations. The composite profile for primarily lead smelting was used to capture industrial lead and zinc emissions. Measurement uncertainties were provided with the composite lead smelting profile. In the absence of specific information, uncertainties for the steel and copper profiles were set to 50% of the mass fraction of each species.

CMB was run with and without the source elimination option, which eliminates negative source contributions from the model solution. If the source elimination option is selected, the source with the largest negative contribution is removed from the calculation, and another fit is attempted. This process is repeated until there are no sources with negative contributions. When the source elimination option was not used for the STL-SS dataset, average $PM_{2.5}$ contributions from diesel vehicles (-0.58 μg/m^3), industrial lead and zinc (-0.08 μg/m^3), and secondary organic carbon (-2.05 μg/m^3) were negative. Thus, despite resulting in a slightly higher average chi-squared value (4.66 v. 2.63), the source elimination option was determined to provide more realistic estimates of source contributions. The source elimination option also allowed CMB to better reconstruct $PM_{2.5}$ mass, with an average modeled-to-measured $PM_{2.5}$ ratio of 90.8% with source elimination compared to 90.0% without source elimination. Therefore, the source elimination option was selected to provide more reasonable source impact estimates. Average source impacts and uncertainties (route mean squared) are show in Figure 5 below.

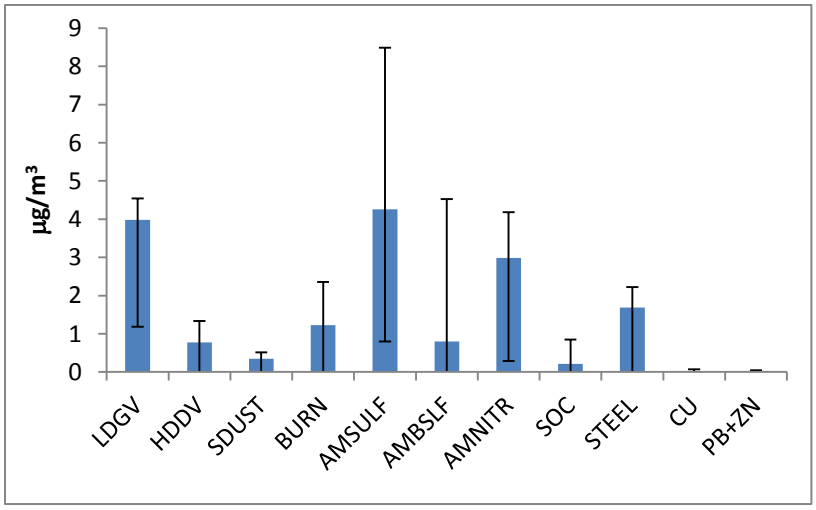


Figure 5: CMB Results using Source Elimination

One major disadvantage of using CMB (with source elimination) to reconstruct $PM_{2.5}$ mass at the STL-SS is that the fitting algorithm failed to converge for 253 of the 686 days with valid measurement data. This was most likely due to co-linearity between two or more of the fitting sources. Further, the model calculated zero impact from diesel vehicles on an additional 141 days and zero impact from SOC on an additional 336 days, including during the summer.

Chemical Mass Balance - Molecular Markers

Bae et al. used CMB-MM to estimate source contributions to OC concentrations at the STL-SS between June 2001 and May 2003. Calculations were performed in EPA CMB8.2 using the effective variance, least square solution. Ambient concentration data used in the CMB calculation consisted of the 1-in-6 day organic molecular marker, EC, OC, silicon, and aluminum measurements collected at the STL-SS. The emissions source profiles used in the study were obtained from previous source testing (Hildemann, Markowski et al. 1991; Rogge, Hildemann et al. 1993; Rogge, Hildemann et al. 1993; Rogge, Hildemann et al. 1993; Schauer, Kleeman et al. 1999; Schauer, Kleeman et al. 1999; Schauer, Fraser et al. 2002). The study considered the following emissions sources: road dust, vegetative detritus, wood smoke, natural gas combustion, diesel exhaust, spark ignition gasoline vehicle exhaust, and gasoline smoker

exhaust (Bae, Schauer et al. 2006). Bae et al. assumed that all of the OC mass that was not attributed to these primary sources was SOA. Daily, 1-in-6 day contributions to total OC and associated uncertainties were provided in Appendix B of Bae et al. 2006. Model performance statistics, including chi-squared values and modeled-to-measured OC ratios were provided in Appendix A of the paper. Average contributions to OC mass and associated uncertainties are shown in Figure 6. Source contributions to $PM_{2.5}$ were estimated based on the source-specific OC to $PM_{2.5}$ ratios provided in Table 3 of Bae et al. 2006 (see Table 2).

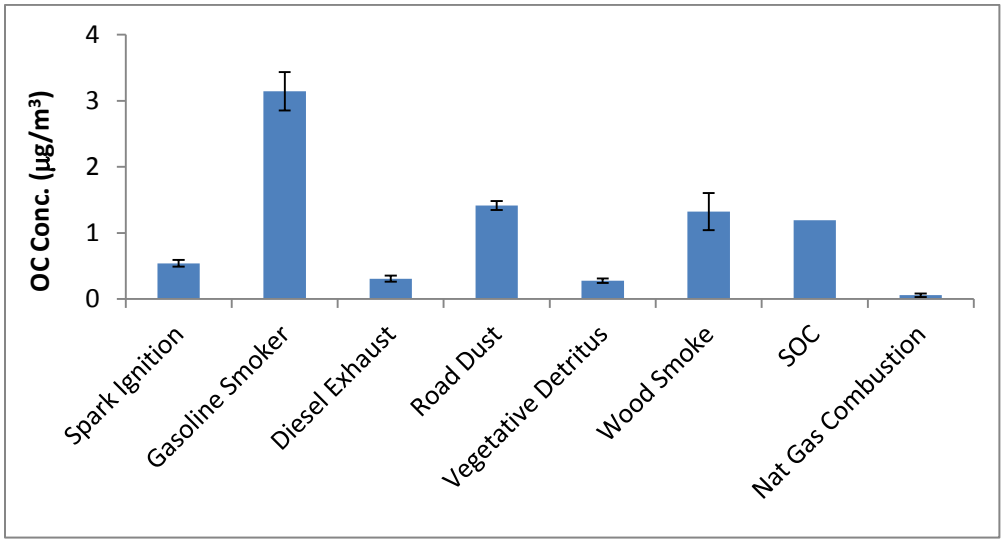


Figure 6: Average OC Contributions and Uncertainties Calculated using CMB-MM

Community Multi-scale Air Quality Model

Baek generated CTM SA results for the entire US using a tracer method within the Community Multiscale Air Quality Model (CMAQ-TR). The CMAQ-TR model used DDM-3D to investigate first order sensitivities of $PM_{2.5}$ concentrations to perturbations in 28 sources of primary aerosol emissions. Baek ran CMAQ-TR for 29 days in July 2001 and 29 days in January 2002, based on emissions and meteorological inputs developed as part of the 2002 VISTAS modeling (Baek 2009). The CMAQ-TR results included estimates of source contributions to major ionic species (i.e., sulfate, nitrate, ammonium), OC, EC, and total $PM_{2.5}$ mass (Baek 2009).

Figure 7 shows the average source contributions for July 2001 and January 2002 calculated using CMAQ-TR. As expected, estimated sulfate concentrations are higher in the summer than in the winter, whereas estimated nitrate concentrations are higher in the winter. Additionally, estimated SOC concentrations are also higher in July 2001 than in January 2002, when photochemistry is more significant. Conversely, estimated gasoline vehicle, dust, and burn impacts are higher in the winter than in the summer. Metal processing and diesel vehicle impacts are similar between the two months, with slightly higher estimated concentrations in January 2002 than in July 2001.

Modeled-to-measured $PM_{2.5}$ ratios were calculated to evaluate model performance. Daily modeled $PM_{2.5}$ concentrations were calculated as the sum of the estimated contributions from the nine sources

considered for the ensemble. The nine CMAQ-TR sources considered for the ensemble reconstructed 76.45 $\pm$ 26.34% of the $PM_{2.5}$ mass in July 2001 and 117.65 $\pm$ 33.52% of the $PM_{2.5}$ mass in January 2002.

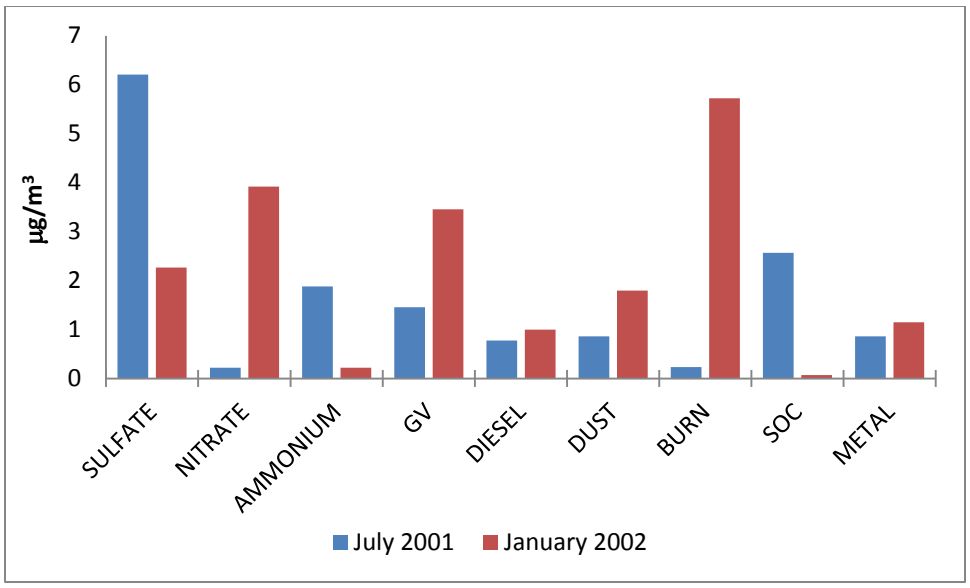


Figure 7: Average Source Contributions for July 2001 and January 2002 using CMAQ-TR

Ensemble-Trained Source Apportionment

Ensemble-average source impacts for July 2001 and January 2002 at the STL-SS were calculated from the PMF, PMF-MM, CMB, CMB-MM, and CMAQ-TR results described above. This study considered nine categories of $PM_{2.5}$ emissions: gasoline vehicle exhaust and resuspended road dust (GV), diesel vehicle exhaust (DV), (DUST), biomass burning (BURN), metals processing (METAL), secondary organic carbon (SOC), secondary sulfate (SULFATE), secondary nitrate (NITRATE), and ammonium (AMMONIUM). These nine source categories incorporate about 70% of the inventoried primary $PM_{2.5}$ sources in St. Louis and the dominate sources that contribute to secondary pollutant formation. The various emissions sources considered in the five different SA approaches were binned into these 9 source categories as show in Table 2. Daily contributions of secondary ionic species (i.e., SULFATE, NITRATE, and AMMONIUM) to $PM_{2.5}$ mass were estimated using only the PMF results. The secondary sulfate and nitrate factors accounted for the majority of the modeled sulfate (85%), nitrate (80%), and ammonium (81%) mass in PMF. However, there was some bleeding of these secondary species into primary factors (i.e., gasoline vehicles) that were not anticipated to contain these species. Thus, the secondary sulfate, nitrate, and ammonium contributions were calculated by summing the predicted mass of these species across all factors. Sulfate was, however, left in the steel processing factor, since the MBSP indicates that a significant mass fraction of composite steel processing emissions are sulfate (40%). Although PMF did not resolve a SOC factor, the OC in the secondary sulfate and nitrate factors was assumed to be from secondary sources.

Since PMF-MM and CMB-MM calculate source contributions to OC rather than $PM_{2.5}$ mass, source contributions from these methods were scaled by source specific OC to $PM_{2.5}$ ratios provided in the literature (Table 2). Since PMF-MM did not resolve separate gasoline and diesel vehicle factors, the

combined mobile source factor was split into a gasoline and a diesel component using the ratio of gasoline to vehicle impacts predicted by CMAQ-TR.

Table 2: Source Specific OC to $PM_{2.5}$ Ratios used for the Molecular Marker Results

Source Apportionment Model	Source Category	OC to $PM_{2.5}$ Ratio	Citation
CMB-MM	Spark Ignition	0.27	(Lough, Schauer et al. 2005; Bae, Schauer et al. 2006)
	Gasoline Smoker	0.58	(Lough, Schauer et al. 2005; Bae, Schauer et al. 2006)
	Road Dust	0.14	(Hildemann, Markowski et al. 1991; Bae, Schauer et al. 2006)
	Diesel Exhaust	0.30	(Schauer, Kleeman et al. 1999; Bae, Schauer et al. 2006)
	Vegetative Detritus	0.32	(Hildemann, Markowski et al. 1991; Bae, Schauer et al. 2006)
	Wood Smoke	0.56	(Schauer, Kleeman et al. 2001; Bae, Schauer et al. 2006)
	Secondary Organic Aerosol	0.58	(Bae, Schauer et al. 2006)
	Natural Gas Combustion	0.85	(Rogge, Hildemann et al. 1993; Bae, Schauer et al. 2006)
PMF-MM	Diesel Vehicles	0.23	(Zheng, Cass et al. 2007; USEPA 2010)
	Gasoline Vehicles	0.68	(Zheng, Cass et al. 2007; USEPA 2010)
	Dust	0.49	(Chow, Watson et al. 2004; Zheng, Cass et al. 2007)
	Biomass Burning	0.70	(Schauer, Kleeman et al. 2001; Chow, Watson et al. 2004; Zheng, Cass et al. 2007)

Ensemble-average source impacts were then calculated using Equation 8 for two one month periods: July 2001 and January 2002. In this work, each SA method was given an equal weighting in the ensemble calculation. Source impact estimates were not available for every model for every day in July 2001 and January 2002. To prevent the ensemble-average source contributions from being biased by the availability (or unavailability) of model results on a particular day, source contributions were normalized by the global mean source impact across all methods (i.e., $S_{kj} - \langle S_{kj} \rangle$). Ensemble calculations were then performed on these normalized values and the average source contributions were rescaled by the global mean source impacts.

Ensemble-average source impacts for July 2001 and January 2002 were then used to calculate optimized winter and summer EBSPs. Optimized source profiles were calculated using a chemical mass balance approach (Equation 1), which minimized daily chi-squared values (Equation 2). Calculations were preformed in Microsoft Excel, using a non-linear solver package developed by Frontline Systems (Frontline Systems 1991-2011). During each optimization, the source profiles were initially set to the

MBSPs, and were constrained to the MBSP +/- 3*σ_{MBSP} as long as these values were between 0 and 1. Additionally, the total mass fraction of the species in the source profiles were restricted to values less than or equal to one. This calculation specified an organic matter to organic carbon ratio of 1.2 for GV, DV, DUST, and METAL and an organic matter to organic carbon ratio of 1.4 for BURN. OC/EC ratios for GV and DV were constrained to values between 0.80 and 4.0 and 0.17 and 1.25, respectively (Marmur, Mulholland et al. 2007). The OC/EC ratio for the BURN profile was constrained to values greater than or equal to 3. Lastly, the total carbon fraction (OC + EC) in the GV, DV, and BURN profiles were constrained to values greater than or equal to 0.5.

As expected, the optimized source profiles varied from day-to-day within the provided constraints. Daily source profiles were averaged for July 2001 and January 2002 to produce the summer and winter EBSPs. The uncertainties associated with the EBSPs were calculated from the standard deviation of the daily optimized source profiles. The EBSPs and associated uncertainties were then used to calculate daily source contributions using the EPA CMB software. The summer EBSPs were applied to daily measurements collected between April and October, while the winter EBSPs were applied to daily measurement collected between November and March.

Table 3: Source Categories used in the Ensemble

Ensemble Source Categories	CMAQ-TR Sources	CMB Sources	CMB-MM Sources	PMF Sources	PMF-MM Sources
GV	• On-road gasoline • Aircraft • Gasoline engine - leisure craft • Paved road dust • Unpaved road dust • Non-road gasoline	• Gasoline vehicle exhaust	• Spark ignition gasoline exhaust • Gasoline smoker exhaust • Road dust	• Gasoline vehicle exhaust	• Mobile factor (split using GV/DV ratios from other methods)
DV	• On-road diesel • Non-road diesel	• Diesel vehicle exhaust	• Diesel exhaust	• Diesel vehicle exhaust	• Mobile factor (split using GV/DV ratios from other methods)
DUST	• Other fugitive dust • Construction dust	• Dust	• Vegetative detritus	• Resuspended soil 1 • Resuspended soil 2	• Resuspended soil
BURN	• Wood/bark industrial combustion • Agricultural burning • Wildfire • Prescribed fire • Fireplaces • Yard waste burning	• Biomass burning	• Wood smoke	• Biomass burning	• Wood combustion
METAL	• Metal industry	• Steel processing • Primary lead smelting (industrial lead and zinc) • Copper processing		• Industrial copper • Industrial lead • Industrial zinc • Steel processing	
SOC		• Secondary organic carbon	• Other OC		• Secondary organic aerosol
Unassigned	• Meat cooking		• Natural gas combustion		• Winter combustion 1 • Winter combustion 2 • Point source 1 • Point source 2

Ensemble Source Categories	CMAQ-TR Sources	CMB Sources	CMB-MM Sources	PMF Sources	PMF-MM Sources
	• Distillate oil combustion • Pulp, paper and wood processing • Cement kilns • Mineral industrial process • Petroleum and solvent evaporation • Other industrial process • Natural gas - other • Natural gas - residential heating • Coal burning				

Chapter 4: Results

Comparison of Source Contributions Estimated by the Five Preliminary SA Models

An inter-comparison of the source impacts estimated by CMB, PMF, CMAQ-TR, CMB-MM, and PMF-MM was preformed prior to incorporating these results in the ensemble-trained SA methodology. While consistency among the different methods does not necessarily indicate accurate results, this comparison provides insight into the biases and uncertainties associated with specific techniques. Pair wise correlation coefficients were calculated between each of the five methods, for the gasoline vehicle, diesel vehicle, dust, biomass burning, metal processing, and SOC source categories. Dust and biomass burning sources are the only two source categories for which the impacts estimated using two or more of the five SA techniques had correlation coefficients above 0.7. Figure 8 shows the five source/method pairs with the highest correlation coefficients. For the dust source, source impacts calculated in PMF v. CMB and PMF v. PMF-MM are associated with intercepts close to zero ($0.14 \pm 4.1 \times 10^{-5}$ and -0.49 ± 0.029); however, impacts estimated by PMF v. CMB have a slope less than one ($0.31 \pm 1.7 \times 10^{-5}$), whereas impacts estimated by PMF v. PMF-MM have a slope greater than one (1.6 ± 0.038). In comparison, dust impacts estimated in CMB v. PMF-MM are associated with an intercept that was significantly less than zero (-1.1 ± 0.032) and a slope greater than one (4.0 ± 0.065). For the biomass burning source, impacts calculated in CMB-MM v. PMF-MM have an intercept close to zero (0.18 ± 0.025) and a slope a little greater than one (1.52 ± 0.097). Since CMAQ results are only available for July 2001 and January 2001, and PMF-MM results are only available for one in every six days during the time period of interest, the regression analysis between these two methods was only based on nine data points. Thus, both the slope (-0.082 ± 0.37) and intercept (4.1 ± 1.2) of the biomass burning impacts predict by CMAQ v. PMF-MM are associated with relatively high standard errors.

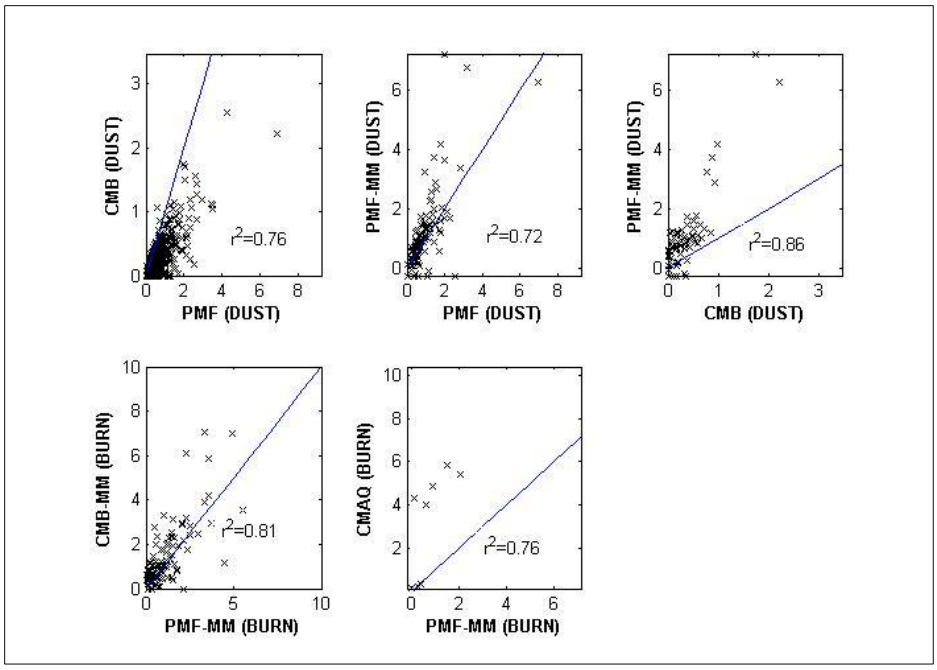


Figure 8: Correlation Analysis of Source Impacts Predicted by CMB, PMF, CMB-MM, PMF-MM, and CMAQ-TR

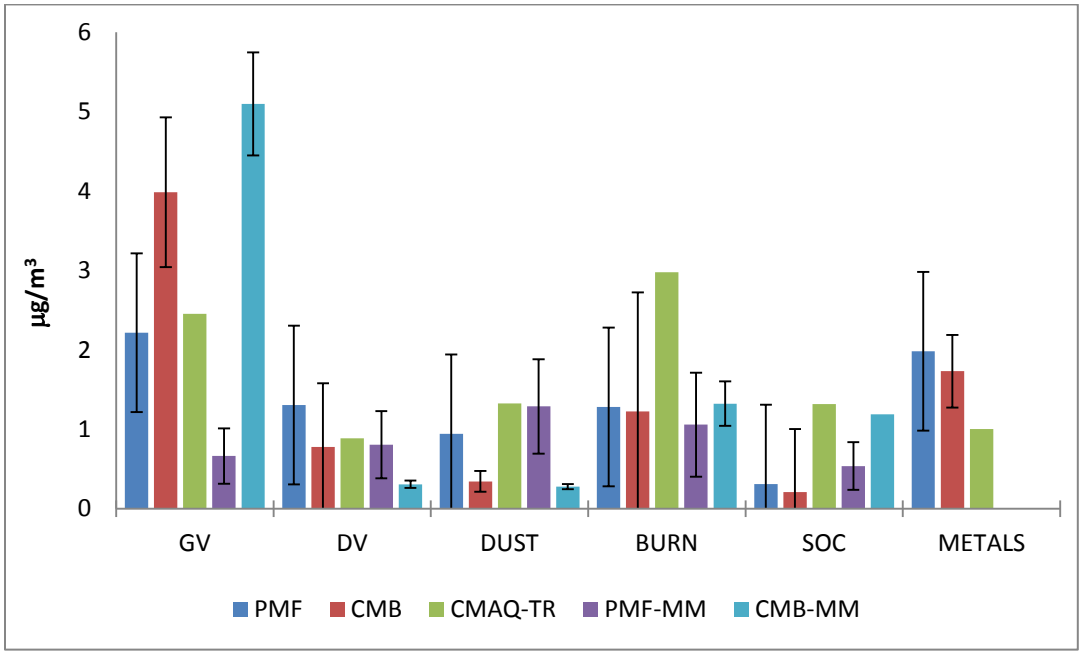


Figure 9: Comparison of $PM_{2.5}$ Source Impacts Estimated by CMB, PMF, CMB-MM, PMF-MM, and CMAQ-TR

A comparison of the average contributions predicted by the five methods shows that CMB, PMF, and CMAQ-TR predicted similar $PM_{2.5}$ impacts from secondary sulfate, nitrate, and ammonium. These results suggest that secondary sulfate, nitrate, and ammonium comprise between 47 and 49% of the average $PM_{2.5}$ mass at the STL-SS. Figure 9 shows average $PM_{2.5}$ contributions estimated by CMB, PMF, CMB-MM, PMF-MM and CMAQ-TR for the five primary sources and SOC. The two CMB methods predict the highest average gasoline vehicle impacts, at 4.0 $\mu g/m^3$ for CMB and 5.1 $\mu g/m^3$ for CMB-MM.

Gasoline vehicles account for 24% of the $PM_{2.5}$ mass reconstructed in CMB, as opposed to 14% in PMF and 13% in CMAQ-TR (Figure 9). Conversely, PMF (1.3 $\mu g/m^3$) predicts the highest average diesel vehicle impacts. As such, diesel vehicles account for 8% of the $PM_{2.5}$ mass reconstructed by PMF, compared with 5% for both CMB and CMAQ-TR. Alternatively, CMAQ-TR predicts the highest dust and burn impacts. Dust comprises 7% of the $PM_{2.5}$ modeled in CMAQ-TR, compared to 2% in CMB and 6% in PMF. This is a known bias in the current CMAQ SMOKE system. CMAQ-TR calculated that biomass burning impacts account for 16% of the average $PM_{2.5}$ impacts, compared to 8% for both CMB and PMF. Conversely, CMAQ-TR and CMB-MM predict the highest average SOC impacts. Lastly, CMB and PMF calculate higher average metals processing impacts than CMAQ-TR. CMB estimates higher steel impacts than PMF, while PMF estimates higher industrial lead, copper, and zinc impacts.

As discussed previously, PMF reconstructs 99.9% (σ=12.5%) of the measured $PM_{2.5}$ mass, whereas CMB reconstructs 90.8% (σ=14.8%) of the measured $PM_{2.5}$ mass. For CMAQ-TR, the nine source categories considered in the ensemble method account for 76.5% (σ=26.3%) of the measured $PM_{2.5}$ mass in July 2001 and 117.7% (σ=33.5%) of the measured $PM_{2.5}$ mass in January 2002. Alternatively, PMF-MM reconstructs 98.4% (σ=40.6%) of the average OC mass. Conversely, the seven primary sources considered in Bae et al.'s CMB-MM analysis (road dust, vegetative detritus, wood smoke, natural gas combustion, diesel exhaust, spark ignition gasoline vehicle exhaust, and gasoline smoker exhaust) reconstruct 84.3% (σ=47.6%) of the average OC mass. Since, Bae et al. 2006 assumed that any OC mass not attributed to the primary sources was SOC, OC contributions from these primary and secondary sources always account for 100% of the OC mass.

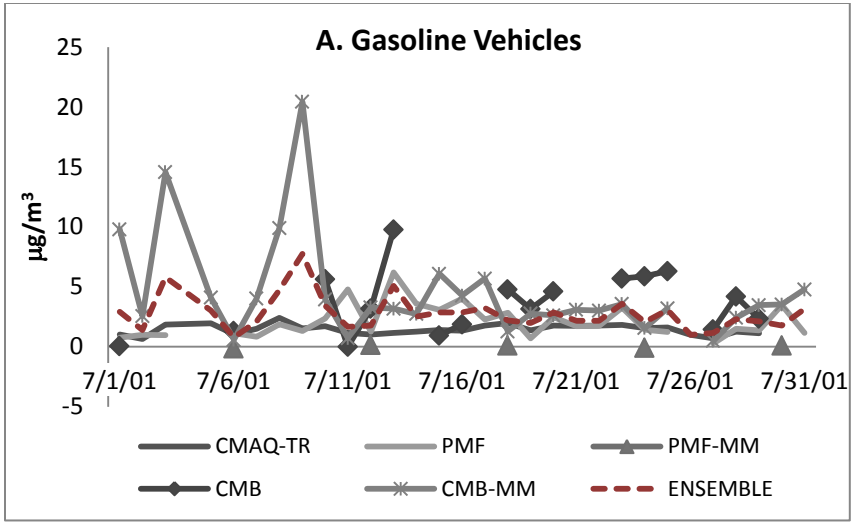
Ensemble Results

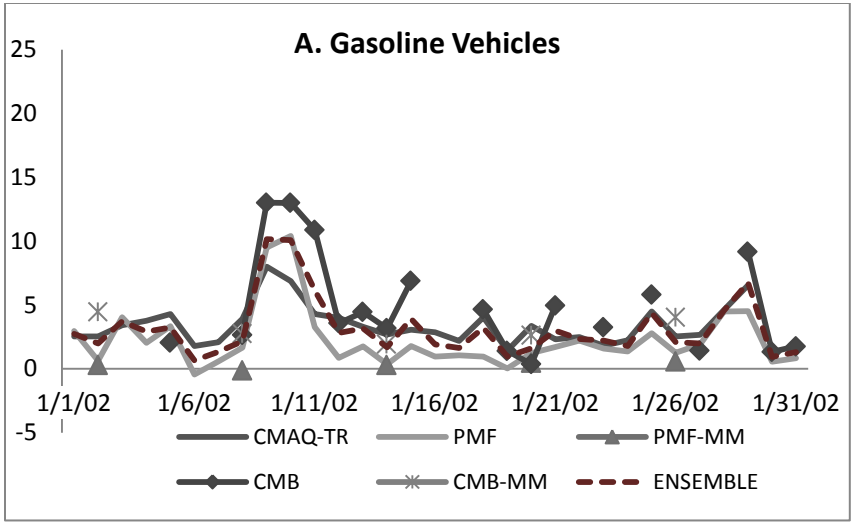
Ensemble-Average Source Contributions for July 2001 and January 2002

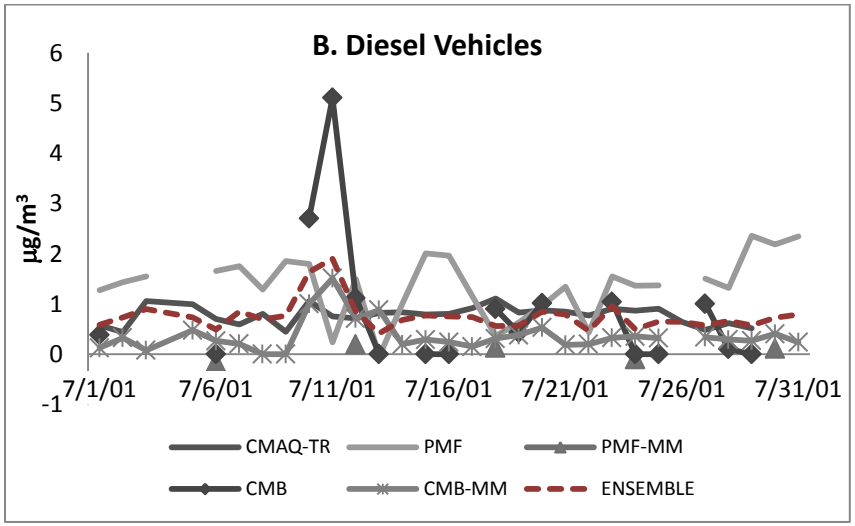
Source impacts estimated by PMF, CMB, PMF-MM, CMB-MM, and CMAQ-TR were combined to calculate ensemble-average source impacts for July 2001 and January 2002. Ensemble-averaged source impacts generally have reduced day-to-day variability compared to PMF, CMB, PMF-MM, and CMB-MM, as expected (Figure). For example, CMB-MM calculates spikes in gasoline vehicle impacts on 7/3/2001 (14.6 $\mu g/m^3$) and 7/9/2001 (20.4 $\mu g/m^3$). While the ensemble method also shows slightly elevated gasoline vehicle impacts on these two days, impacts were reduced to 5.8 and 7.6 $\mu g/m^3$ on 7/3/2001 and 7/9/2001, respectively. A similar dampening effect is observed for the diesel vehicles, dust, and biomass burning sources in July. Source impacts calculated from the five different SA methods exhibit less day-to-day variability in January 2002 than in July 2001 (Figure). As anticipated, the CMAQ-TR results exhibit less day-to-day variability than the other receptor-based approaches. Conversely, the ensemble results for the composite metals processing source exhibit the day-to-day variability expected from a point source or combination of point sources. The four receptor modeling approaches (CMB, PMF, CMB-MM, and PMF-MM) tend to predict elevated metals processing impacts on the same days in July and January (e.g., July 10th and 18th and January 9th, 18th, 21st, and 29th).

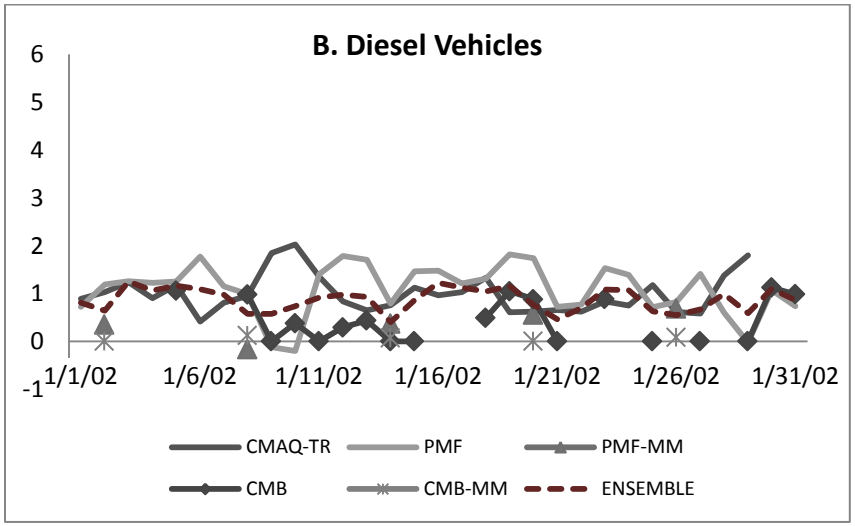
In addition to reducing day-to-day variability in the source impacts, the ensemble method also reduces the number of zero impact days and days without valid model results. Since the two molecular marker models are based on 1-in-6 day particle-bound organic measurements, PMF-MM and CMB-MM results are only available for a limited number of days during the sample period. Additionally, the CMB model fails to converge to provide source impacts on more than 35% of the 686 days with complete measurement data, including a number of days in July 2001 and January 2002. Further, CMB estimates that a number of significant sources, including diesel vehicles and SOC, do not impact $PM_{2.5}$ concentrations on numerous additional days. Since the ensemble method uses information from a number of different methods, source impacts can be resolved on days when CMB or the molecular markers methods do not provide source impacts.

Lastly, ensemble-average source impacts were calculated for a summer month and a winter month to consider seasonal variations in source impacts and source compositions. A comparison of the source impacts estimated for July 2001 and January 2002 demonstrates that calculated SOC concentrations are higher in July than in January. Conversely, estimated biomass burning impacts are higher in January than in July. The other sources do not appear to exhibit a discernible seasonable trend.


A. Gasoline Vehicles
µg/m³
25
20
15
10
5
0
-5
7/1/01
7/6/01
7/11/01
7/16/01
7/21/01
7/26/01
7/31/01
CMAQ-TR
PMF
PMF-MM
CMB
CMB-MM
ENSEMBLE


A. Gasoline Vehicles
25
20
15
10
5
0
-5
1/1/02
1/6/02
1/11/02
1/16/02
1/21/02
1/26/02
1/31/02
CMAQ-TR
PMF
PMF-MM
CMB
CMB-MM
ENSEMBLE


B. Diesel Vehicles
µg/m³
6
5
4
3
2
1
0
-1
7/1/01
7/6/01
7/11/01
7/16/01
7/21/01
7/26/01
7/31/01
CMAQ-TR
PMF
PMF-MM
CMB
CMB-MM
ENSEMBLE


B. Diesel Vehicles
6
5
4
3
2
1
0
-1
1/1/02
1/6/02
1/11/02
1/16/02
1/21/02
1/26/02
1/31/02
CMAQ-TR
PMF
PMF-MM
CMB
CMB-MM
ENSEMBLE

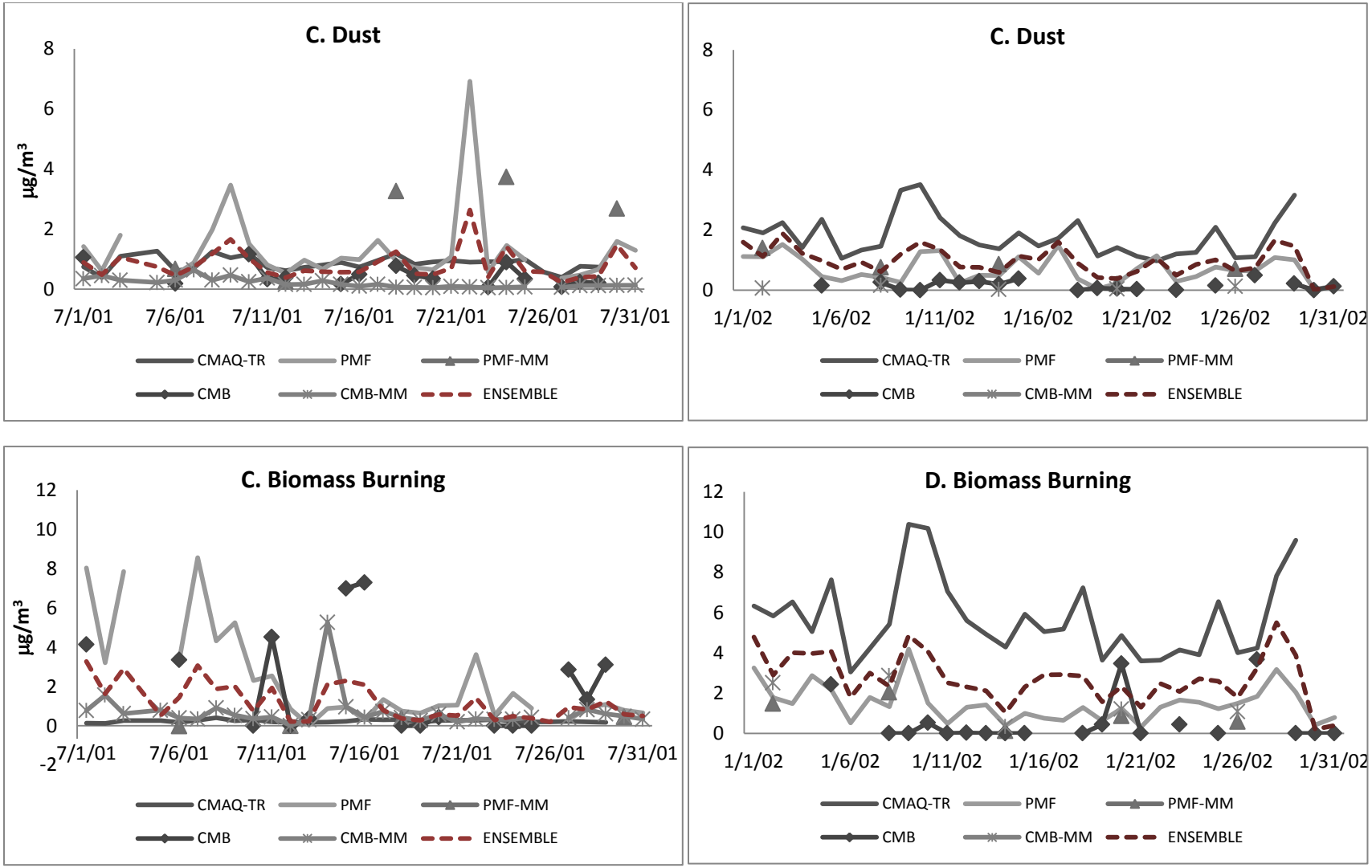
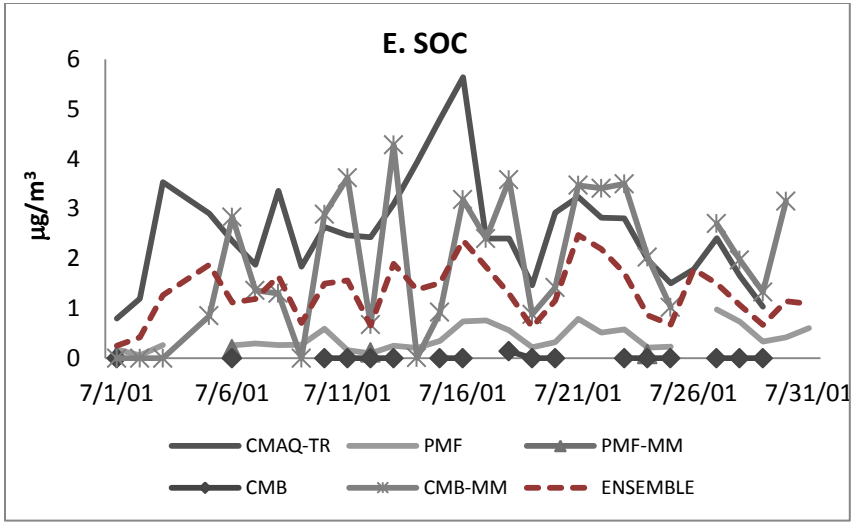
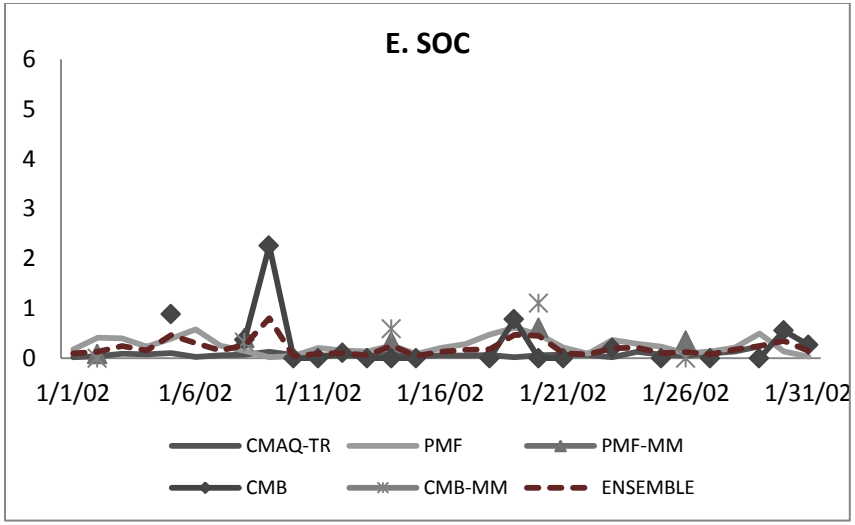
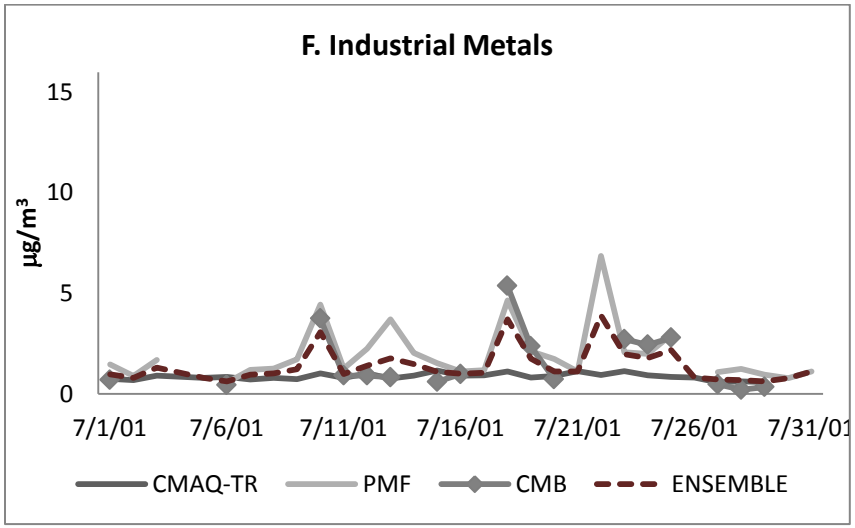


Figure 10: Ensemble-Average Source Contributions for July 2001 and January 2002







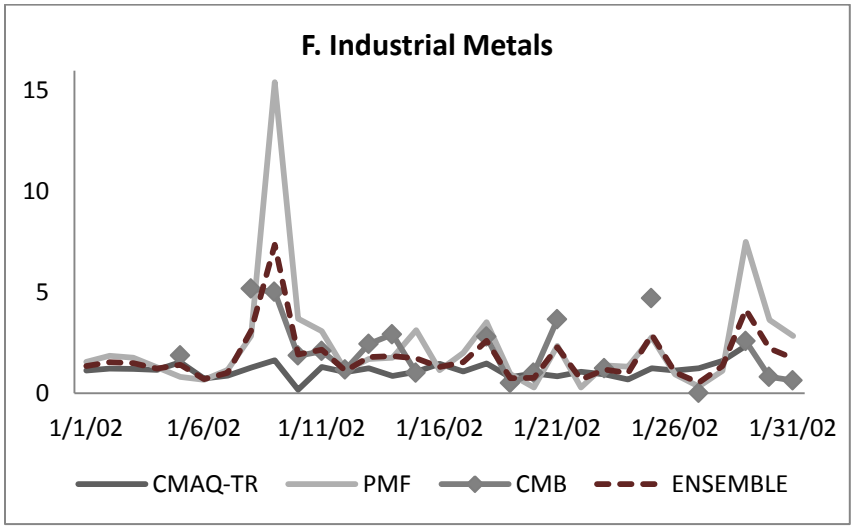


Figure 10 Continued

Ensemble-Based Source Profile Compositions

A CMB-based method was used to calculate summer and winter EBSPs using the ensemble-averaged source impacts (Figure 11). As expected, the source profiles that minimized the daily chi-squared values vary slightly from day-to-day. The optimized summer EBSPs were calculated as the average of the daily profiles for July 2001, whereas the winter EBSPs were calculated as the average of daily profiles for January 2002. Source profile uncertainties for the winter and summer were calculated as the standard deviation of the daily July and January profiles, respectively. This procedure generally produces EBSPs with lower source profile uncertainties than the MBSPs (Figure 11). A comparison of the optimized summer and winter EBSPs to the MBSPs shows greater inter-profile variability in the weight percentages of the XRF species than the ionic species (i.e., sulfate, nitrate, and ammonium), EC, and OC. Thus, the EC/OC ratios and total OC and EC fractions for the gasoline and diesel vehicle sources are nearly identical for the MBSPs and EBSPs. OC/EC ratios for the gasoline source are 0.429 , 0.432, and 0.428 for the MBSP, summer EBSP, and winter EBSP, respectively. EC/OC ratios for the diesel source are 3.71 for the MBSP and the winter EBSP and 3.72 for the summer EBSP.

Figure 11 shows a slight reduction in the relative quantities of silicon, aluminum, and potassium in the dust profile. The mass fractions of silicon, aluminum, and potassium are 0.27, 0.095, and 0.0092 for the MBSP, compared to 0.25, 0.093, and 0.0069 for the summer EBSP and 0.22, 0.092, and 0.00060 for the winter EBSP. The winter and summer biomass burning EBSPs also shows a slight reduction in potassium content (0.052 and 0.026 for summer and winter, respectively) compared to the MBSP (0.057). These reductions may be due in part to decreases in the contributions of the 16 fitting species used in the ensemble analysis to the total profile mass. For the MBSPs, the 16 fitting species account for 92% of the mass of the dust source and 119% of the biomass burning source. Conversely, the summer EBSPs account for 88.7% of the mass of dust source and 99.9% of the biomass burning source. The winter EBSPs account for even less of the mass of these two sources, reconstructing 80% of the dust source and 96% of the biomass burning source.

The composite metals processing source shows the greatest difference between the EBSPs and the MBSP. Since the metals processing source is most likely a composite of point sources and is by nature difficult to characterize, the composition of this profile was expected to exhibit more day-to-day variability than the other non-point sources and potentially differ from the composite MBSPs obtained from Speciate. For example, the EBSPs contain substantially less arsenic, silicon, and zinc than the composite metals processing MBSP. As with the other sources, the 16 fitting species used in the EBSPs account for less of the total mass of the metals processing source than the MBSP; however, the difference in the reconstructed mass was more significant. The optimized winter and summer EBSPs only capture 33% and 28% of the metal processing profile mass, respectively, compared to the 71% characterized by the composite metals processing MBSP. In general, the EBSPs account for a smaller fraction of emissions than the MBSPs. This result is likely caused by the constraints, which provided an upper bound mass fraction for each profile, but no lower bound.

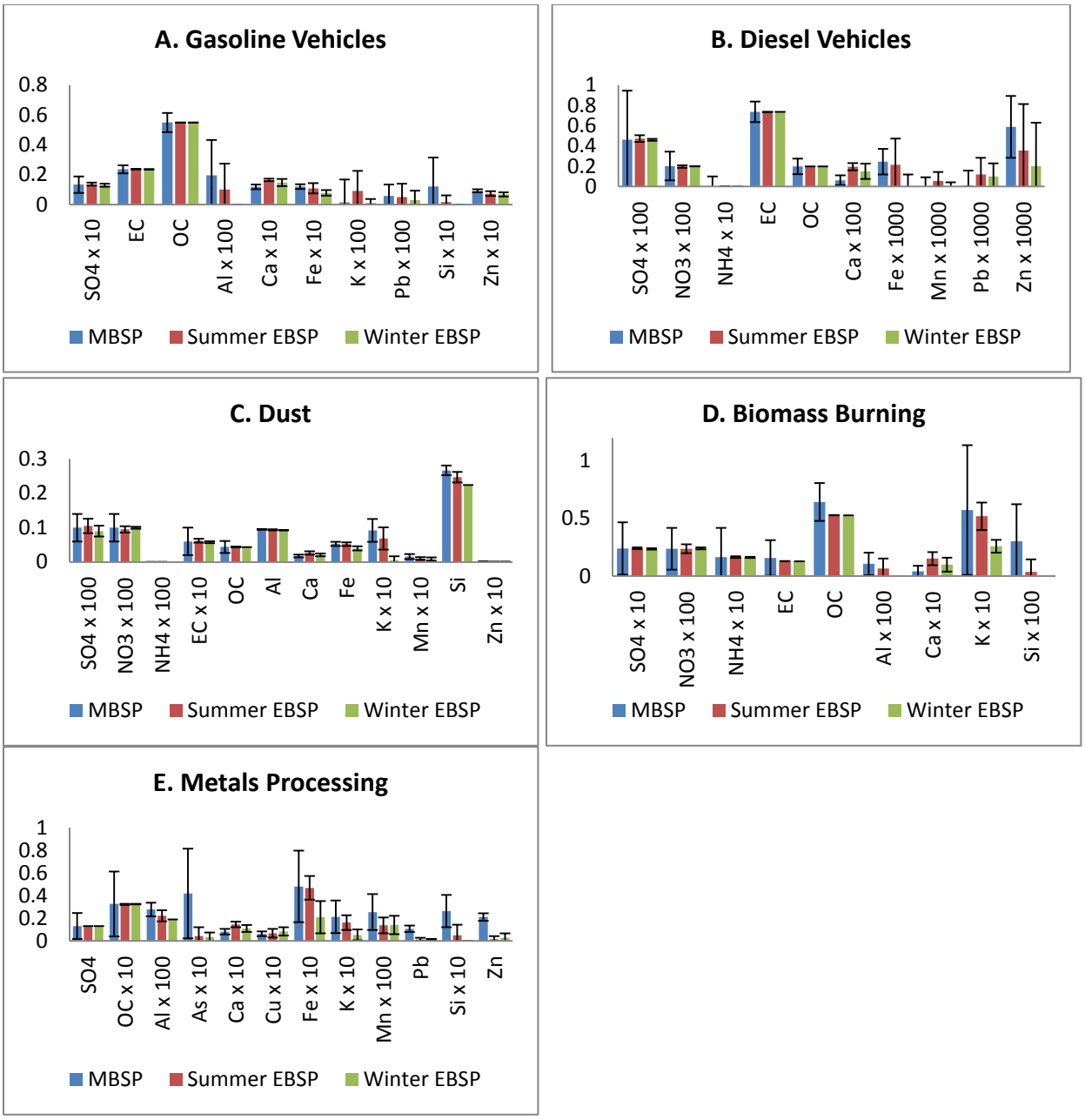


Figure 11: Comparison of Winter and Summer EBSPs to the MBSPs

Source Apportionment Using the EBSPs

EPA'S CMB software was used to calculate revised source impacts using the winter and summer EBSPs. Based on the initial CMB results, CMB was run using the source elimination option, which precluded negative source impacts. Daily source impacts at the STL-SS between June 2001 and May 2003 estimated using CMB and the EBSPs were compared to source impacts estimated using the five original SA models. For the gasoline vehicle source category, the source impacts predicted by PMF and CMB are most correlated with the source impacts estimated using the EBSPS, both with r^2 values of 0.66. Diesel vehicle impacts calculated using the MBSPs in CMB are most correlated with impacts calculated using the EBSPs, with an r^2 value of 0.85. Dust impacts estimated by PMF, CMB, and PMF-MM are all fairly

well-correlated with dust impacts estimated using CMB and the EBSPs, with r^2 values of 0.77, 0.84, and 0.82 respectively. Burn impacts calculated using PMF (r^2=0.63) and SOC impacts calculated using CMB (r^2=0.67) are better correlated with the ensemble results than the other methods. Lastly, metals processing impacts estimated by PMF and CMB are better correlated with metals processing impacts estimated using the ensemble method than with metals processing impacts estimated using CMAQ-TR, with r^2 values of 0.72 and 0.71 compared with 0.22. This result was expected, since the average emissions inventory data used in CMAQ-TR is not anticipated to capture the day-to-day variability in this emission source category.

Comparison of CMB using the EBSPs to CMB using the MBSPs

Unlike the initial CMB results, which were incorporated into the ensemble method, the EBSPs use a single, composite metals processing source to represent industrial copper, lead, zinc, and steel emissions. Thus, CMB results using the EBSPs were compared to CMB results using a set of MBSPs, which included a composite, measurement-based metals processing profile. The composite, measurement-based metals processing profile was calculated as the weighted average of the three measurement-based metals processing profiles used in the initial CMB analysis (copper processing, steel processing, and lead smelting) based on the average contributions calculated in CMB. Source profile uncertainties were calculated from the uncertainties in the individual metals profiles using propagation of errors. The ensemble-based source contributions were compared to source contributions obtained using the composite metals processing MBSP (Table 4). The composite metals processing MBSP was also used as a starting point when calculating the metals processing EBSPs.

Table 4: Comparison of CMB Results using EBSPs and MBSPs

	MBSPs[a]	EBSPs	
χ^2_{CMB}[b]	31.87±72.33	2.84±2.98	
% Mass[c]	87.8±13.6	94.7±15.4	
	Average Contributions, μg/m³		
	MBSPs	**EBSPs**	**r^2**
GV	3.57±0.60	3.55±0.86	0.66
DV	0.56±0.58	0.91±0.36	0.85
DUST	0.67±0.13	0.43±0.08	0.84
BURN	2.21±1.22	1.20±0.50	0.53
SOC	0.14±0.53	0.57±0.45	0.67
METAL	0.17±0.11	2.01±0.60	0.71

[a] Composite metals processing profile

[b] Chi-squared value output by CMB; considers both measurement and source profile uncertainty

[c] Ratio of modeled-to-measured $PM_{2.5}$ mass

Source impacts estimated using the MBSPs with the composite metal processing profile were fairly well-correlated with source impacts estimated using the EBSPs (Table 4). A comparison of the average source impacts estimated using the two methods indicated that the EBSPs led to an increase in predicted diesel vehicle and SOC impacts, and a decrease in dust and burn impacts. Impacts from the metals processing source are also increased using the EBSPs (from 1% to 12% of the reconstructed

mass). Additionally, the chi-squared value, which reflects the difference between the observed and estimated $PM_{2.5}$ mass, is decreased from 31.9 to 2.8 using the EBSPs rather than the MBSPs (composite metals processing profile). The large standard deviation in the chi-squared value for the MBSPs (composite metals processing profile) indicates that there were some days when the error in the modeled $PM_{2.5}$ is relatively large. Further, the percentage of modeled-to-measured $PM_{2.5}$ mass is closer to 100% with the EBSPs than with the MBSPs (composite metals processing profile).

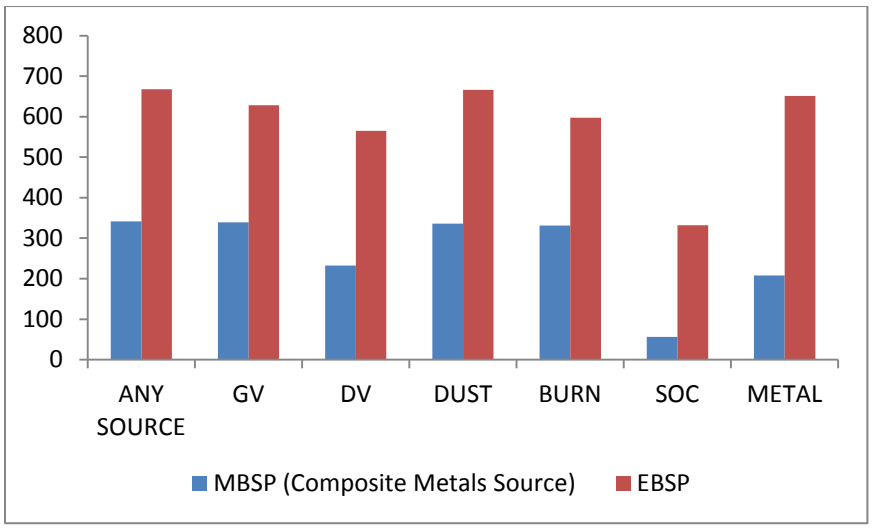


Figure 12: Days with Valid Non-Zero Source Impacts

As discussed previously, one of the limitations of using CMB on the STL-SS dataset is that there were a substantial number of days for which the CMB fitting algorithm failed to converge to a feasible solution, likely due to co-linearity between two or more source profiles. Further, CMB was run using the source elimination option, which selectively eliminates emissions sources until a solution is found where all estimated source impacts are positive. This can lead to one or more sources having no predicted impact on $PM_{2.5}$ concentration for a given day. The CMB fitting algorithm was only able to resolve source impacts for 341 of the 686 days with valid measurement data using the MBSPs (composite metals processing profile) (Figure 12). Further, CMB estimated that diesel vehicles and SOC did not contribute to $PM_{2.5}$ mass on an additional 109 and 285 days, respectively. Using the EBSPs the number of days for which CMB was able to resolve source impacts increases to 668, and the number of days with valid, non-zero source impacts for diesel vehicles and SOC also increases significantly. However, even with the EBSPs, CMB still predicts that diesel vehicles and SOC do not impact $PM_{2.5}$ concentrations on 103 and 336 days, respectively. This reduction in zero impact days explains, in part, the observed increase in gasoline vehicle and SOC impacts observed using the EBSPs instead of the MBSPs (composite metals processing profile). Additionally, the EBSPs increase the number of days with valid, non-zero source impacts for the composite metals source from 208 to 651. However, the metals processing source is assumed to be comprised of a number of point sources, which are only anticipated to hit the monitoring site under certain conditions. Thus, this result combined with the substantial increase in estimated metals processing impacts, could indicate that the EBSPs are causing CMB to apportion too large a

percentage of the $PM_{2.5}$ mass to the composite metals source. Alternatively, since there are numerous metals processing facilities surrounding the STL-SS, it is also possible that industrial metals impact the STL-SS on most days during the time period of interest.

Comparison of CMB using the EBSPs to PMF

Source impact predicted using the EBSPs in CMB (CMB-EBSPs) were also compared to the original PMF results. Although the CMB model outputs chi-squared values, chi-squared values were recalculated using Equation 2 to allow the CMB results to be compared with the PMF results. The PMF results are associated with a lower average chi-squared value than the ensemble results (7400 v. 9930). Additionally, PMF better reconstructs the measured $PM_{2.5}$ mass, with a modeled-to-measured $PM_{2.5}$ ratio of 99.9% $\pm$ 11.3% compared to 94.7% $\pm$ 15.4% for the ensemble method. PMF and the ensemble method predict similar average metals processing and burn impacts, with r^2 values of 0.72 and 0.63 for metals processing and biomass burning, respectively (Table 5). The gasoline vehicle and dust sources predicted by the two methods are relatively well-correlated (r^2 values 0.66 and 0.77); however differences in the average impacts predicted by the two methods indicate that CMB using the EBSPs predicts higher GV impacts than PMF, whereas PMF predicts higher dust impacts than CMB using the EBSPs. Diesel vehicle and SOC impacts are not well-correlated between the two methods.

Table 5: Comparison of the Results of PMF to CMB using the EBSPs

	PMF	CMB-EBSPs	
χ^2	7400$\pm$30600	9930$\pm$26500	
% Mass[b]	99.9$\pm$11.3	94.7$\pm$15.4	
	Average Contributions, $\mu g/m^3$		
	PMF	**CMB-EBSPs**	**r^2**
GV	2.22$\pm$1.51	3.55$\pm$0.86	0.66
DV	1.30$\pm$1.02	0.91$\pm$0.36	0.21
DUST	0.94$\pm$0.59	0.43$\pm$0.08	0.77
BURN	1.28$\pm$0.61	1.20$\pm$0.50	0.63
SOC	0.31$\pm$0.09	0.57$\pm$0.45	0.20
METAL	1.98$\pm$0.91	2.01$\pm$0.60	0.72

[a] Composite metals processing profile
[b] Ratio of modeled-to-measured PM2.5 mass

Sensitivity of CMB Source Apportionment Results to Metals Processing Profiles

One of the greatest challenges associated with SA at the STL-SS is adequately characterizing the metals processing sources impacting this location. In this analysis, MBSPs were developed from source testing conducted on a number of metals processing emissions sources (i.e., copper processing, steel processing, primary lead smelting), which were thought to be similar to those impacting the STL-SS. However, the composition of the particulate matter emitted from individual metals processing facilities may vary widely depending upon the nature of the operations conducted at a specific industrial site (e.g., furnace types, purity of the raw material, emissions controls). Thus, the copper processing, steel processing, and primary lead smelting profiles selected for this analysis may not accurately represent

emissions from the nearby industrial operations affecting the STL-SS. Additionally, the metals processing profiles used in this work were developed from source testing conducted in the mid-to-late 1980s (USEPA 2011). These profiles may not reflect industry-wide operational changes that occurred between the 1980s and the 2001-2003 time period considered for this study.

Factor analytical approaches, such as PMF, and emissions-based SA approaches, such as CMAQ-TR, circumvent the need to accurately characterize individual emissions sources impacting the STL-SS. However, both methods have their own limitations. Due to the time required to run DDM-3D and CMAQ, CMAQ-TR results are only available for two one-month periods in the summer and winter. Further, metals processing impacts in CMAQ are based on nationwide emissions inventory data, which have their own data gaps and uncertainties. Conversely, PMF uses a statistical approach to decompose an ambient concentration data matrix into unique factor profiles and impacts, preventing the need for user input source profiles. However, the user must select an appropriate number of factors and link these factors to individual emission sources or source categories. Further, as was seen in this work, there may be some bleeding of major $PM_{2.5}$ constituents, such sulfate, nitrate, and ammonium, into sources that are not typically thought to contain these compounds. Also, PMF can predict zero impact days for sources that likely impacted the site to a significant degree.

The use of EBSPs rather than MBSPs may address some of the issues associated with characterizing the metals processing sources affecting the STL-SS. The EBSPs were optimized within set constraints (MBSP $\pm$ $3\sigma_{MBSP}$) using the ensemble-average source impacts and the chi-squared formulation (Equation 2) as the objective function. In this way, the EBSPs should represent a better estimate of the average composition of the metals processing emissions impacting the STL-SS in July 2001 and January 2002. Additionally, unlike typical CMB applications, use of the summer and winter EBSPs considers the possibility that the composition of particular emissions sources may vary seasonally.

In order to further investigate the sensitivity of the SA results to the composition of the metals processing profile, two separate sets of metals processing profiles were developed from the PMF results. As discussed above, the 11 factor PMF solution resolved four different metals processing factors: steel production, industrial copper, industrial lead, and industrial zinc. The first set of profiles considered each of the four metals processing factors as separate emissions sources (individual PMF metals processing profiles). For the second set of profiles, a composite metals processing profile was developed from the weighted average of the PMF metals factors and average factor contributions (composite PMF metals processing profile). The resulting profiles were used in CMB with the other MBSPs for gasoline vehicles, diesel vehicles, dust, biomass burning, ammonium sulfate, ammonium bisulfate, ammonium nitrate, and SOC. The results were compared with CMB results using individual measurement-based metals processing profiles (lead smelting, steel processing, and copper processing) and a composite measurement-based metals processing profile. Table 6 compares the source impacts predicted using the four different sets of metals processing profiles (individual PMF-based, composite PMF-based, individual measurement-based, and composite measurement-based metals processing profiles) described above with the EBSPs.

The lowest chi-squared values and best mass reconstruction (calculated as the ratio of modeled-to-measured $PM_{2.5}$ mass) were obtained using the EBSPs (Table 6). Conversely, the highest chi-squared value and poorest reconstruction of $PM_{2.5}$ mass were obtained using a composite MBSP to represent industrial metals emissions. Based on a comparison of SA results using the measurement-based and PMF-derived metals profiles, it would appear that multiple metals profiles were generally associated with lower chi-squared values than a single, composite metals processing profile. This is expected because incorporating more sources allows greater flexible in the solution and increases the degrees of freedom. Additionally, the industrial metals emissions impacting the STL-SS likely emanate from a number of different point sources. As such, the composition of this source category is expected to vary considerably from day-to-day, depending on meteorological conditions, operating conditions at nearby industrial facilities, etc. Thus, this source may be best represented by a number of different sources, rather than a single composite source. However, the optimized EBSPs, which produce the lowest chi-squared value, use a composite metals processing profile. This could indicate that the metals processing sources are better characterized by the EBSP than the PMF- or measurement-based source profiles.

Table 6: Comparison of Fitting Statistics and Average Source Contributions to $PM_{2.5}$ Mass using PMF-, Measurement-, and Ensemble-Based Metals Processing Profiles

	MBSPs & Individual Measurement-Based Metals Profiles	MBSPs & Composite Measurement-Based Metals Profile	MBSPs & Individual PMF Metals Profiles	MBSPs & Composite PMF Metals Profile	EBSPs
Chi-Squared	4.66±5.66	31.87±72.33	2.86±7.22	6.15±5.71	2.84±2.98
%Mass	90.81±14.84	87.75±13.60	88.56±12.93	90.03±14.33	94.67±15.37
GV (μg/m^3)	3.98±0.56	3.57±0.60	2.51±0.94	2.36±0.82	3.55±0.86
DV (μg/m^3)	0.78±0.56	0.56±0.58	0.66±0.80	0.61±0.88	0.91±0.36
DUST (μg/m^3)	0.34±0.17	0.67±0.13	0.27±0.13	0.51±0.12	0.43±0.08
BURN (μg/m^3)	1.23±1.13	2.21±1.22	2.42±1.50	2.84±1.54	1.20±0.50
SOC (μg/m^3)	0.21±0.63	0.14±0.53	0.13±0.79	0.13±0.70	0.57±0.45
STEEL (μg/m^3)	1.69±0.53	-	1.08±0.40	-	-
CU (μg/m^3)	0.02±0.05	-	0.26±0.08	-	-
PB (μg/m^3)	-	-	0.31±0.14	-	-
ZN (μg/m^3)	-	-	0.17±0.13	-	-
PB+ZN (μg/m^3)	0.02±0.02	-	0.47±0.19	-	-
METAL (μg/m^3)	1.73±0.54	0.17±0.11	1.81±0.46	1.67±0.47	2.01±0.60

Average source impacts were somewhat sensitive to the selection of the metals processing profile(s) (Table 6). The individual measurement-based metals profiles are associated with the highest average gasoline vehicle impacts, and the measurement-based and ensemble-based metals profiles result in gasoline vehicles contributing more to the total $PM_{2.5}$ mass than the other primary sources. Conversely, SA results for the PMF-based metals profiles indicate that burn impacts are similar to or greater than gasoline vehicles impacts. Further, SA results using the MBSPs with the composite measurement-based metals profile show the highest dust impacts and the lowest metals processing impacts. On the other

hand, the EBSPs yield the highest estimate of average SOC and metals processing impacts. Diesel vehicle impacts estimated using the five sets of profiles describe above were better correlated than the other sources, with an average r^2 value of 0.88. Conversely, the SOC contributions estimated using the five different sets of metals processing profiles were associated with the lowest average r^2 value (0.54). Lastly, source impacts calculated using the EBSPs were less correlated with the other methods than the other methods were with each other.

CMB was able to resolve source impacts for the greatest fraction of the 686 days with complete measurement data using the EBSPs (Figure 13). The EBSPs also decreased the number of zero impact days for several major emissions sources, including vehicles and SOC. Reductions in the number of zero impacts days for the composite metals source are also seen with the EBSPs. Conversely, CMB was able to resolve source impacts for the fewest number of days with the individual PMF-derived metals profiles (238 out of 686 days).

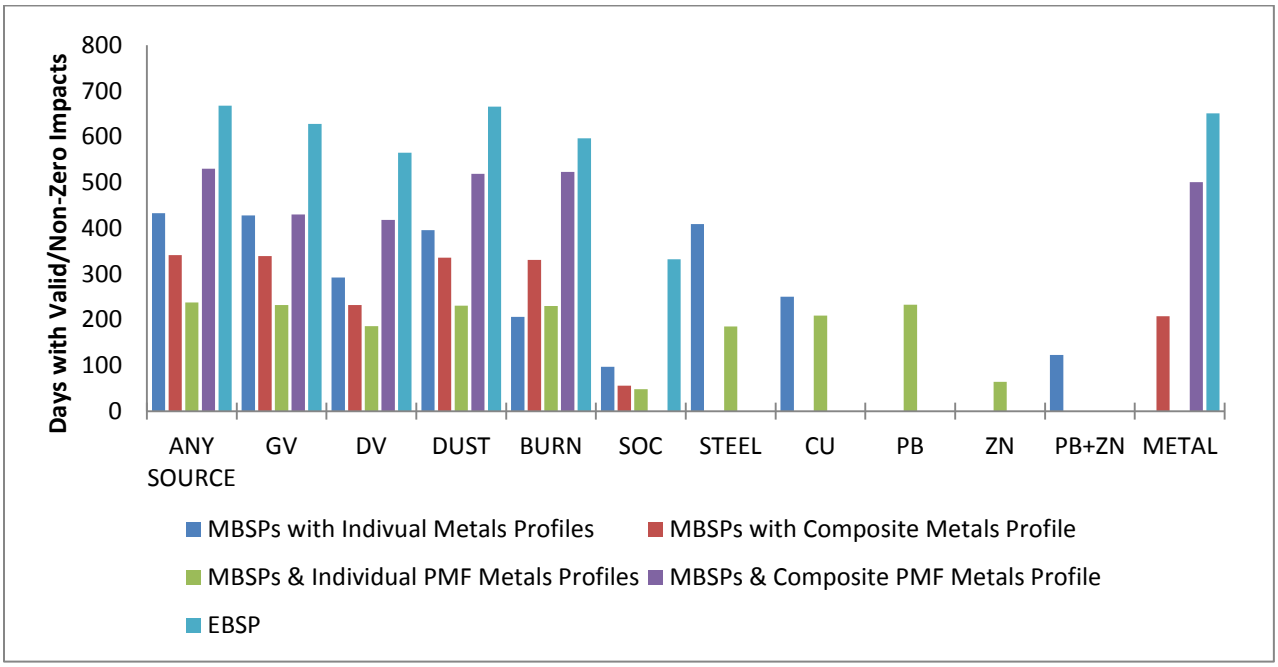


Figure 13: Comparison of Days with Valid, Non-Zero Source Impacts using PMF-, Measurement-, and Ensemble-Based Metals Processing Profiles

Chapter 5: Discussion and Conclusions

This work compared five different techniques for SA at the STL-SS between June 2001 and May 2003: CMB, PMF, CMAQ-TR, CMB-MM, and PMF-MM. A comparison of the average source impacts estimated by these methods indicated that CMB-MM and CMB calculated the highest average gasoline vehicle impacts, whereas PMF calculated the highest diesel vehicles impacts. Conversely, CMAQ-TR calculated the highest average dust and burn impacts, while the highest SOC impacts were calculated by CMB-MM and CMAQ-TR. Lastly, PMF calculated the highest impacts from industrial metals.

Impacts from these five methods were then averaged for two one month periods in July 2001 and January 2002 when CMAQ-TR data was available. Ensemble-average source impacts for July 2001 and January 2002 were then used to calculate summer and winter EBSPs. The EBSPs were generally similar to the MBSPs, with the exception of the metals processing source, which was not believed to be well-characterized by the MBSPs.

SA using the EBSPs offered several distinct advantages over SA using the MBSPs. Primarily, CMB is able to calculate source impacts on more days using the EBSPs than the MBSPs. The EBSPs also reduce zero impact days for major emissions sources, such as diesel vehicles and SOC. Further, goodness of fit statistics, such as the chi-squared value and ratio of modeled-to-measured $PM_{2.5}$ mass, are improved using the EBSPs. Conversely, the EBSPs result in a higher estimate of metals processing impacts and the fewest zero impact days for this source. This result could indicate that $PM_{2.5}$ from industrial point sources impacted the STL-SS on most days during the time period of interest. However, this could also indicate that the ensemble-based metals processing source includes other uncharacterized regional sources in addition to the intended industrial point sources. This could cause CMB to apportion too much $PM_{2.5}$ mass to the composite industrial metals source.

When the ensemble results were compared with the original PMF results, PMF was associated with a lower average chi-squared value and better $PM_{2.5}$ mass reconstruction. While the ensemble method estimated higher gasoline vehicle impacts and SOC impacts than PMF, the two methods estimated similar metals processing and burn impacts. Similarities in average metals processing impacts calculated using these two methodologies could indicate that the metals processing impacts estimated by the ensemble method are in fact reasonable.

While the ensemble method offers several advantages over the other CMB-based SA approaches used at the STL-SS, this method is also subject to several limitations. Primarily, results from a number of different SA methods were needed to calculate the EBSPs and ensemble-based source impacts. For the STL-SS, it was possible to incorporate results from previous SA studies into the ensemble calculations. However, it may not be practical to apply the ensemble method to receptor locations where prior SA results are not available, due to the time requirements necessary to run the five different SA models. Additionally, the molecular marker methods used herein require ambient measurements of particle-bound organic molecular makers. Since this data is not collected as part of EPA's $PM_{2.5}$ monitoring networks, molecular markers data is not commonly available. Thus, CMB-MM and PMF-MM may not be applied at all desired receptor locations.

Further, while we believe that the ensemble method outperforms CMB based on goodness of fit statistics and reductions in zero impact days, the true source impacts at the STL-SS cannot be directly measured. Thus, it is impossible to conclusively determine which method is most accurate. However, since the ensemble method incorporates results from a number of different SA strategies, each with its own biases and uncertainties, this method is thought to avoid the pitfalls associated with any single method.

In addition to testing the newly developed ensemble-based SA technique, this work also served to highlight some of the difficulties associated with using traditional SA approaches at a receptor location impacted by multiple point sources. First, the resolution of separate gasoline and diesel vehicle sources in PMF was found to be sensitive to the choice of XRF fitting species. If all 21 of the XRF species considered in Garlock 2006 were used as fitting species in PMF, separate gasoline and diesel vehicles sources were not resolved.[3] However, when 10 of these 21 species were removed from the analysis, due to low detections and/or lack of importance for factor identification, the 11 factor solution revealed separate gasoline and diesel vehicle factors. This highlights the subjectivities associated with the use of factor analytical approaches for SA.

Further, this analysis suggests that the source impacts estimated by CMB can be sensitive to changes in a subset of the source profiles input into the model. As discussed above, one of the greatest challenges associated with applying CMB to the STL-SS was the development of representative metals processing profiles. Since the composition of the metals processing sources impacting the STL-SS were not well-known, several different techniques were used to characterize this source. CMB results using individual and composite metals processing profiles developed from source testing data (MBSPs) and from relevant PMF factors were compared with CMB results using the EBSPs. This analysis suggested that calculated impacts for all emissions sources varied depending upon the characterization of the metal processing source, even though metals processing impacts accounted for less than 15% of the modeled $PM_{2.5}$ mass. This result indicates that subjective decisions, such as the characterization of local point sources, can impact SA results. These results elucidate the benefits of using several different SA models, since there are limitations, subjectivities, and uncertainties associated with each available approach. For example, if the sources impacting a particular receptor location are not well-characterized it may be useful to use both a factor analytical and CMB approach. One allows the user to investigate source impacts at the receptor site without quantitatively characterizing the composition of important emissions sources, while the other avoids the subjectivities associated with selecting an appropriate number of factors and linking these factors to real-world sources. Additionally, while CMAQ-TR results may not be available for the entire time period of interest, these results may be used to understand if the source impacts predicted by receptor models generally agree with the meteorological and emissions inventory data. Conversely, receptor models may better characterize the day-to-day variability in source impacts missed by CTMs, such as CMAQ. Lastly, molecular markers methods provide SA results using an

[3] For a number of species, Garlock used different method detection limits and analytical uncertainties than were used in this work. Efforts to Garlocks's results using the data processing procedures outlined in her thesis were not successful.

independent dataset, which is not biased by trace metals; however, molecular marker data is not widely available and must be scaled by source specific $PM_{2.5}$/OC ratios, adding uncertainty to the analysis.

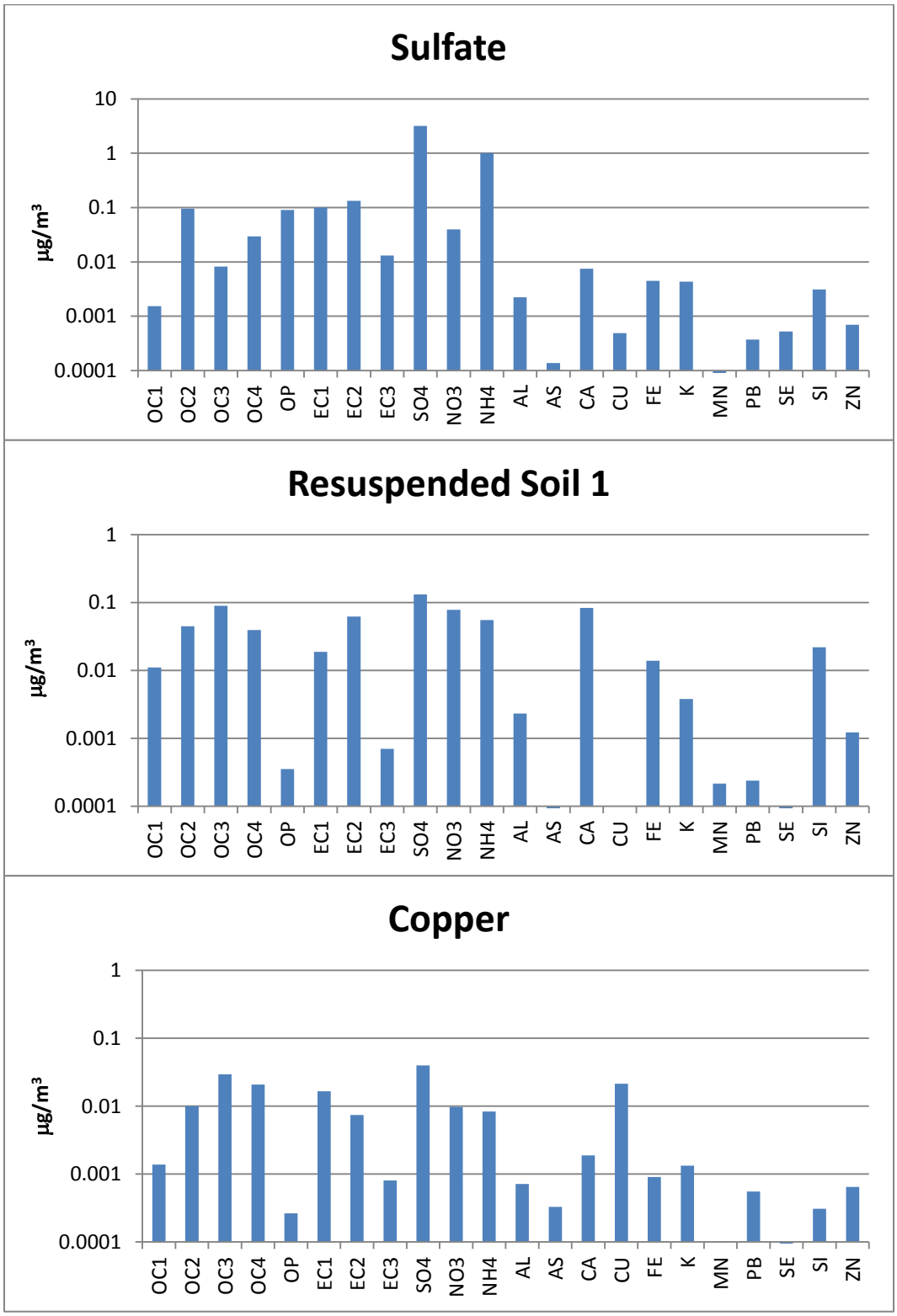
Chapter 6: Future Work

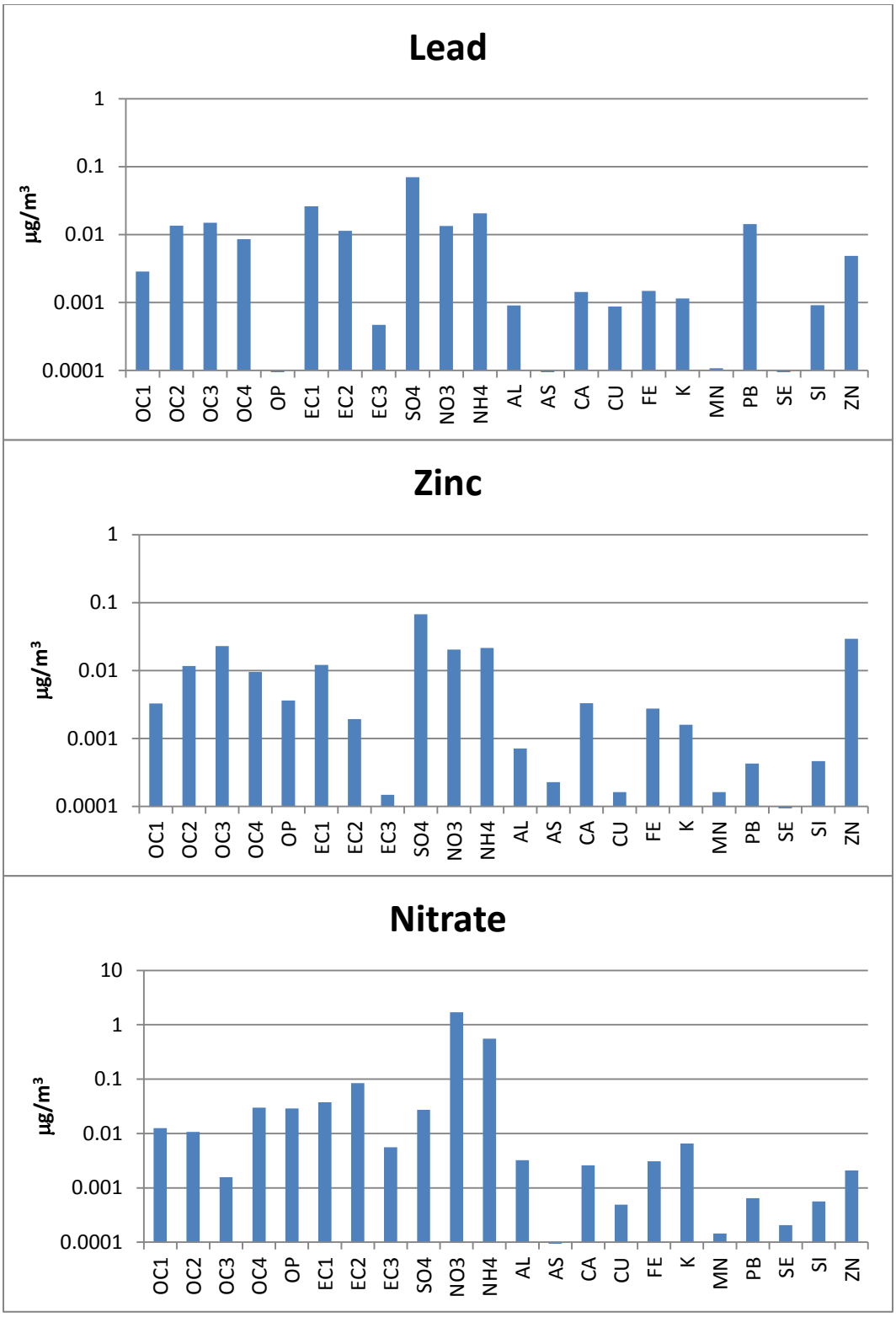
In the future, source impact estimates may be developed for other monitoring locations in St. Louis. There are several EPA monitors located west of the STL-SS in St. Louis, which measure $PM_{2.5}$ components as part of EPA's $PM_{2.5}$ speciation trends network. These monitors include the Blair St. Site (38.656449, -90.198548), the Gratton Cap Site (38.616833, -90.208222), and the Minnesota Avenue Site (38.601778, -90.232778). Measurement data for the Minnesota Avenue Site is available from 2000 through the present, whereas measurements data is available for Gratton Cap Site between 2001 and 2002, and for the Minnesota Avenue Site between 2001 and 2004. SA at these sites would not only provide source impact estimates for a greater geographical area, but also extend the time period for which SA are available in the St. Louis area, since the STL-SS only operated between June 2001 and May 2003.

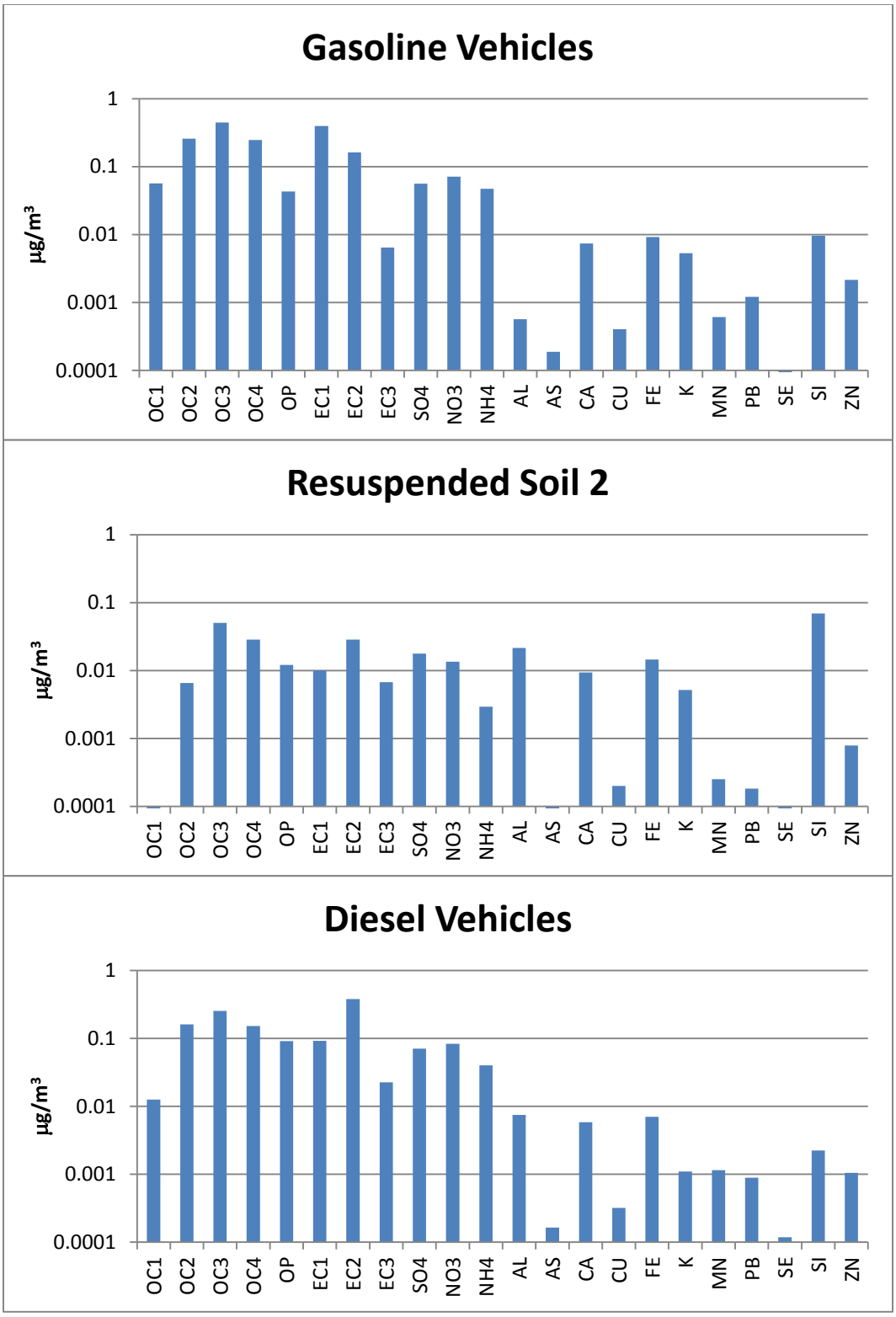
Data restrictions at the three EPA sites preclude or limit the use of several of the methods used at the STL-SS. Since fractionated OC/EC data is not routinely collected at speciation trends network sites, it is unlikely that PMF could be used to resolve gasoline vehicle and diesel vehicle impacts at these sites. Further, particle-bound organics data is also not available for these sites, precluding the use of the two molecular marker methods. Since CMAQ-TR results are only available for July 2001 and January 2002, a long-term SA study at these sites would likely utilize a CMB-based method. Since the EBSPs provided several improvements over the MBSPs for CMB analyses at the STL-SS, it would be informative to evaluate the performance of the STL-SS EBSPs at the nearby EPA monitoring sites.

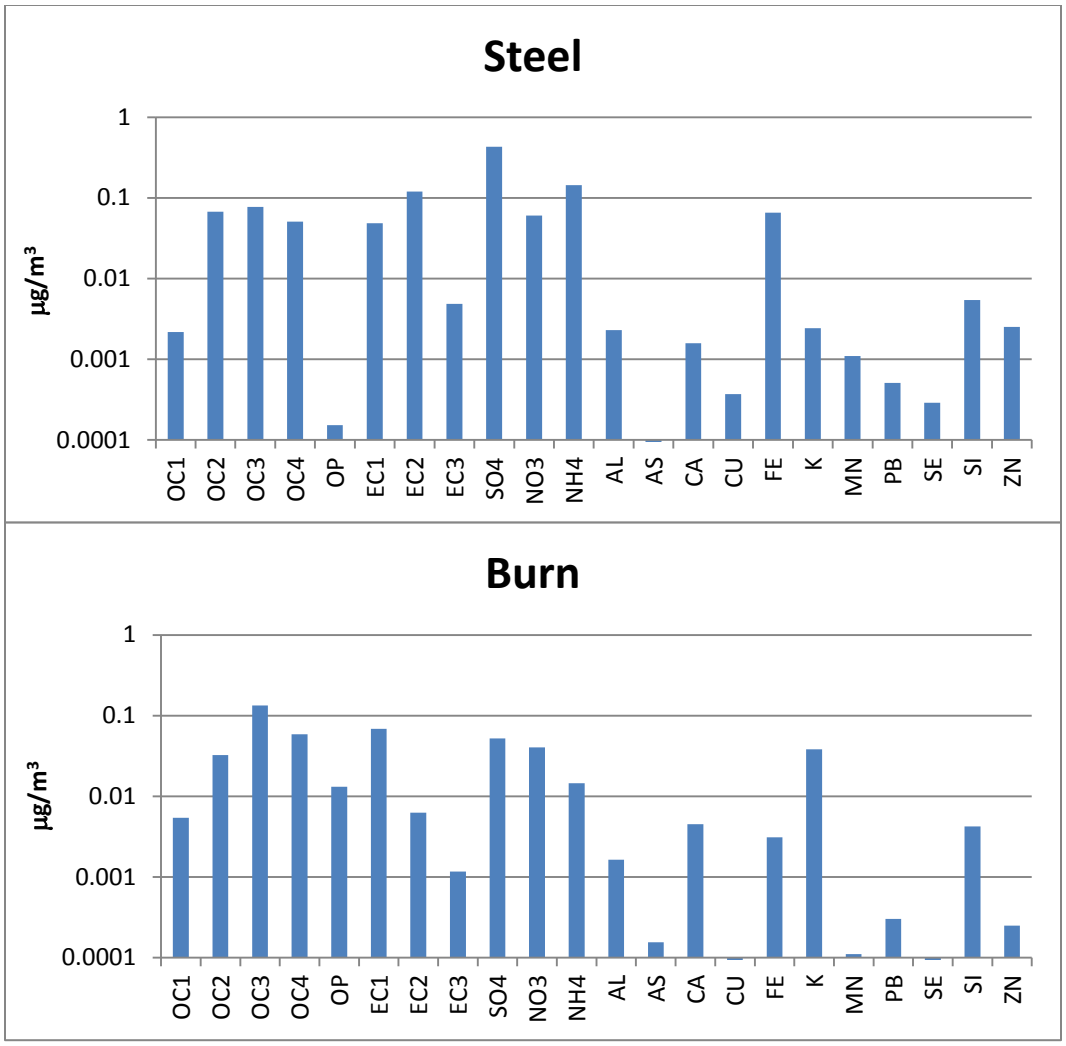
Additionally, the different source impact estimates developed in this work could be used in health studies, linking individual emissions sources to health outcomes, such as cardiorespiratory disease. Based on on-going work at the Jefferson Street site in Atlanta, health associations calculated using the ensemble results may be similar to health associations calculated using the CMB results. However, with the exception of the biomass burning and metal processing sources, the ensemble-results were better correlated with CMB than PMF. Therefore, it may be interesting to compare health associations calculated using the ensemble results to health associations calculated using the PMF results. Further, SOC impacts predicted by the different SA models considered herein were less correlated with one another than estimated impacts from other emission sources. Thus, greater differences in health associations between the different SA methods may be observed for SOC than the other sources.

Appendix A: PMF Source Profiles

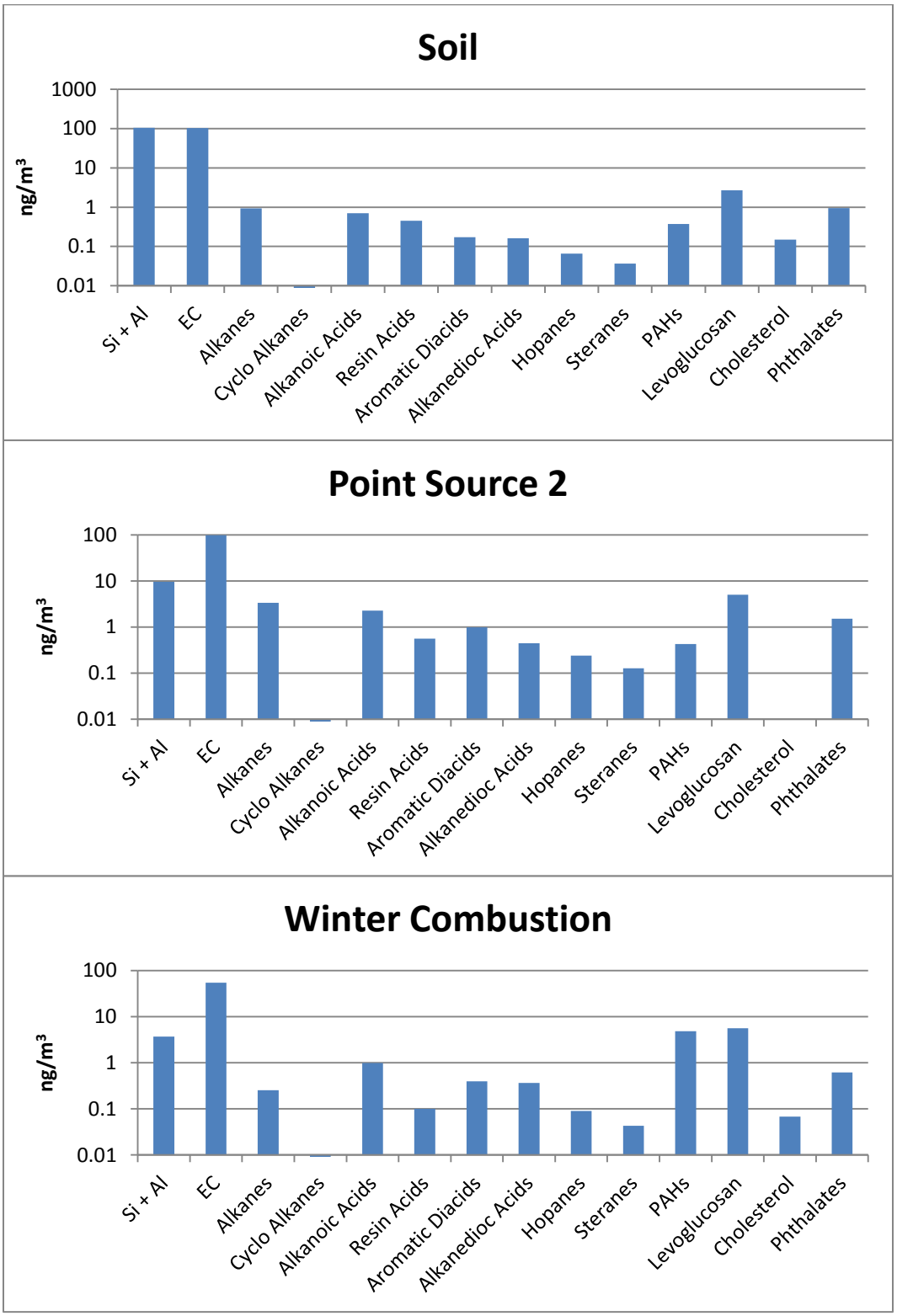


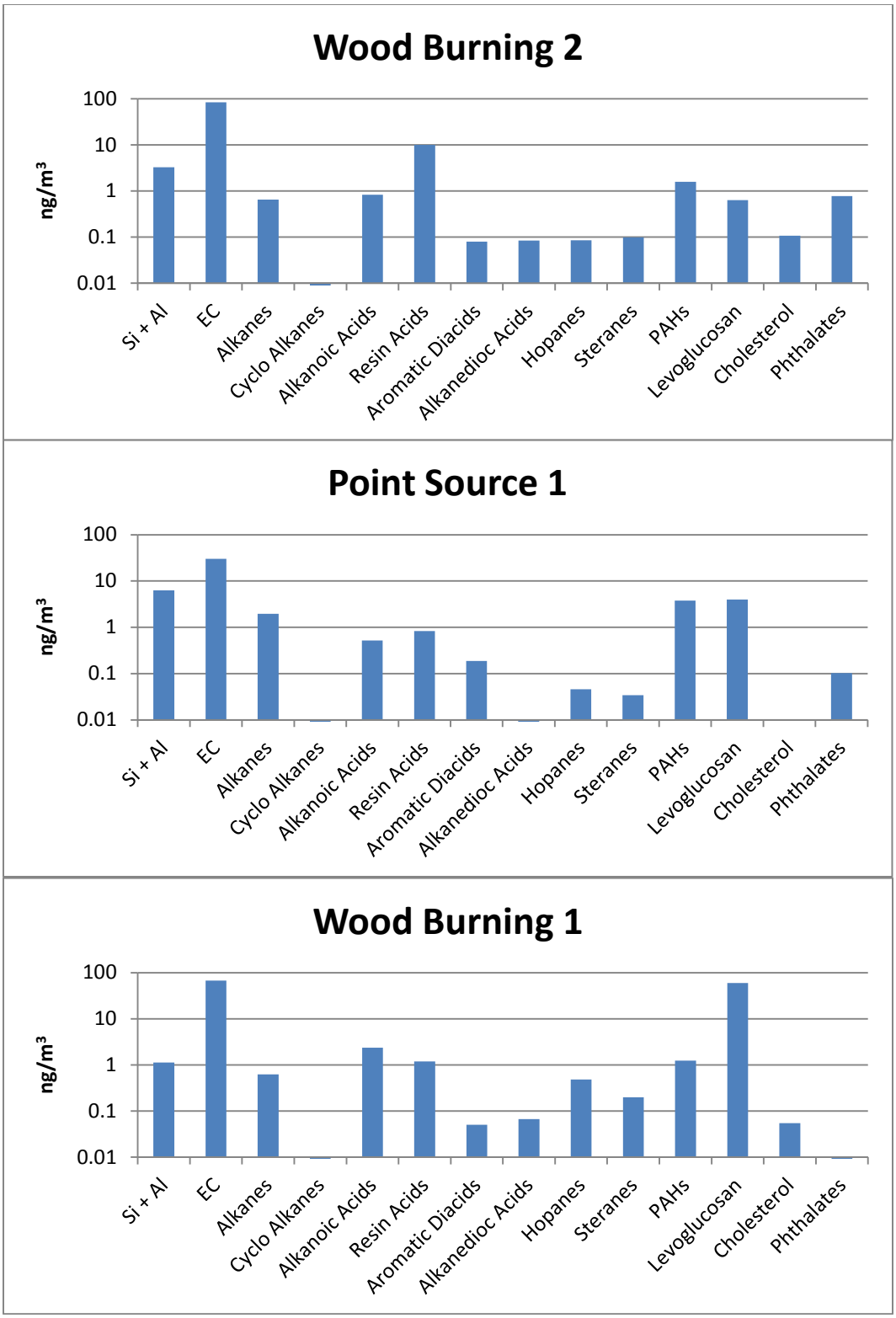

Lead
µg/m³
1
0.1
0.01
0.001
0.0001
OC1 OC2 OC3 OC4 OP EC1 EC2 EC3 SO4 NO3 NH4 AL AS CA CU FE K MN PB SE SI ZN
Zinc
µg/m³
1
0.1
0.01
0.001
0.0001
OC1 OC2 OC3 OC4 OP EC1 EC2 EC3 SO4 NO3 NH4 AL AS CA CU FE K MN PB SE SI ZN
Nitrate
µg/m³
10
1
0.1
0.01
0.001
0.0001
OC1 OC2 OC3 OC4 OP EC1 EC2 EC3 SO4 NO3 NH4 AL AS CA CU FE K MN PB SE SI ZN

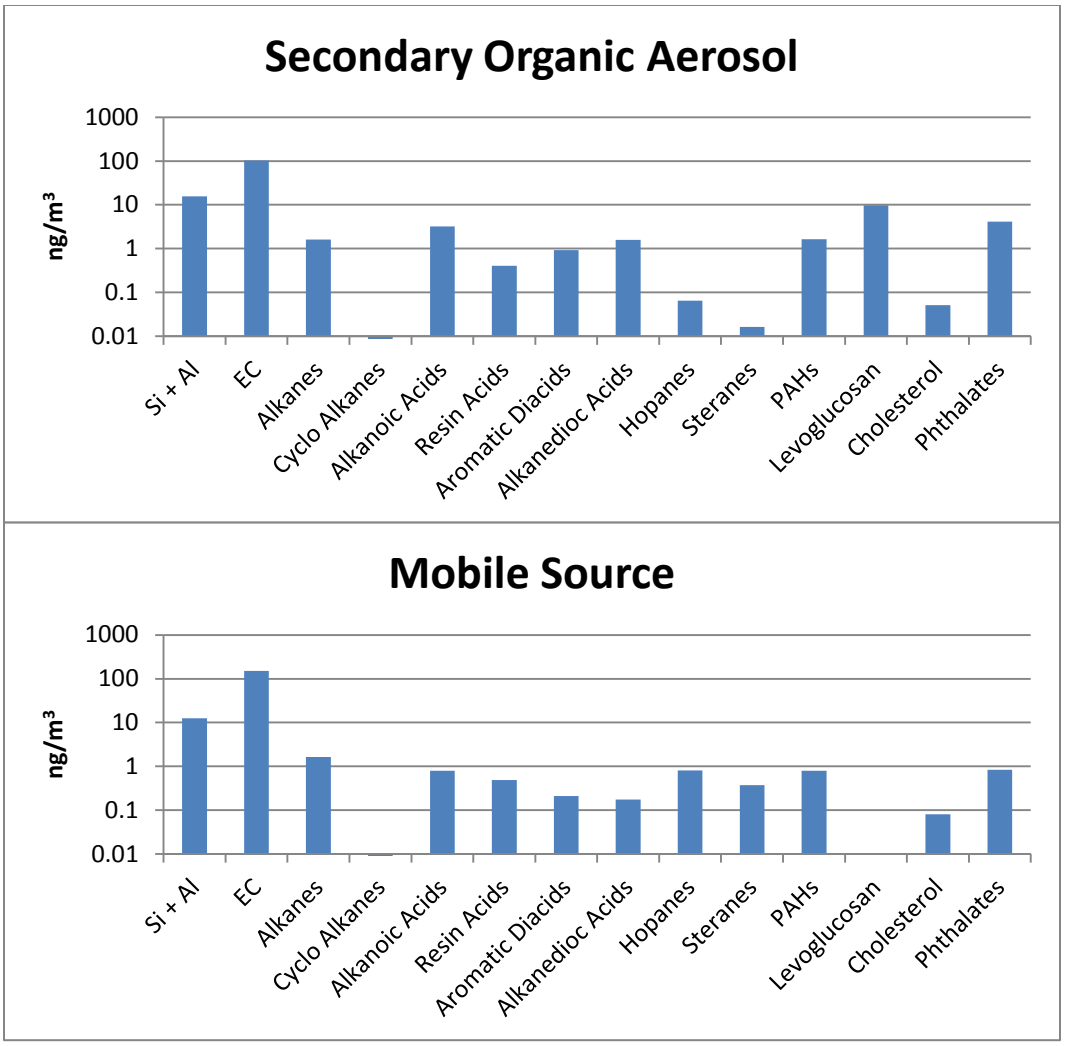

Gasoline Vehicles
µg/m³
1
0.1
0.01
0.001
0.0001
OC1 OC2 OC3 OC4 OP EC1 EC2 EC3 SO4 NO3 NH4 AL AS CA CU FE K MN PB SE SI ZN
Resuspended Soil 2
µg/m³
1
0.1
0.01
0.001
0.0001
OC1 OC2 OC3 OC4 OP EC1 EC2 EC3 SO4 NO3 NH4 AL AS CA CU FE K MN PB SE SI ZN
Diesel Vehicles
µg/m³
1
0.1
0.01
0.001
0.0001
OC1 OC2 OC3 OC4 OP EC1 EC2 EC3 SO4 NO3 NH4 AL AS CA CU FE K MN PB SE SI ZN


Steel
µg/m³
1
0.1
0.01
0.001
0.0001
OC1
OC2
OC3
OC4
OP
EC1
EC2
EC3
SO4
NO3
NH4
AL
AS
CA
CU
FE
K
MN
PB
SE
SI
ZN
Burn
µg/m³
1
0.1
0.01
0.001
0.0001
OC1
OC2
OC3
OC4
OP
EC1
EC2
EC3
SO4
NO3
NH4
AL
AS
CA
CU
FE
K
MN
PB
SE
SI
ZN

Appendix B: PMF-MM Source Profiles




Wood Burning 2
ng/m³
100
10
1
0.1
0.01
Si + Al
EC
Alkanes
Cyclo Alkanes
Alkanoic Acids
Resin Acids
Aromatic Diacids
Alkanedioc Acids
Hopanes
Steranes
PAHs
Levoglucosan
Cholesterol
Phthalates
Point Source 1
ng/m³
100
10
1
0.1
0.01
Si + Al
EC
Alkanes
Cyclo Alkanes
Alkanoic Acids
Resin Acids
Aromatic Diacids
Alkanedioc Acids
Hopanes
Steranes
PAHs
Levoglucosan
Cholesterol
Phthalates
Wood Burning 1
ng/m³
100
10
1
0.1
0.01
Si + Al
EC
Alkanes
Cyclo Alkanes
Alkanoic Acids
Resin Acids
Aromatic Diacids
Alkanedioc Acids
Hopanes
Steranes
PAHs
Levoglucosan
Cholesterol
Phthalates


Secondary Organic Aerosol
ng/m³
1000
100
10
1
0.1
0.01
Si + Al
EC
Alkanes
Cyclo Alkanes
Alkanoic Acids
Resin Acids
Aromatic Diacids
Alkanedioc Acids
Hopanes
Steranes
PAHs
Levoglucosan
Cholesterol
Phthalates
Mobile Source
ng/m³
1000
100
10
1
0.1
0.01
Si + Al
EC
Alkanes
Cyclo Alkanes
Alkanoic Acids
Resin Acids
Aromatic Diacids
Alkanedioc Acids
Hopanes
Steranes
PAHs
Levoglucosan
Cholesterol
Phthalates

References

Atmopsheric Research & Analysis, I. (2003). "Our Studies SEARCH." Retrieved April 4, 2012, 2012, from http://www.atmospheric-research.com/studies/SEARCH/index.html.

Bae, M. S., J. J. Schauer, et al. (2006). "Estimation of the monthly average ratios of organic mass to organic carbon for fine particulate matter at an urban site." Aerosol Science and Technology **40**(12): 1123-1139.

Baek, J. (2009). Improving Aerosol Simulations: Assessing and Improving Emissions and Secondary Organic Aerosol Formation in Air Quality Modeling. Civil and Environmental Engineering. Atlanta, Georgia Institute of Technology. **Doctor of Philosophy:** 202.

Baek, J., S. K. Park, et al. (2005). Source Apportionment of Fine Organic Aerosol Using CMAQ Tracers. The 4th Annual CMAS Models-3 Users' Conference, Durham, NC.

Chow, J. C., J. G. Watson, et al. (2004). "Source profiles for industrial, mobile, and area sources in the Big Bend Regional Aerosol Visibility and Observational study." Chemosphere **54**(2): 185-208.

Coulter, C. T. (2004). EPA-CMB8.2 Users Manual. USEPA. Research Triangle Park, Office of Air Quality Planning & Standards.

Garlock, J. L. (2006). PM2.5 Mass Source Apportionment for the St. Louis - Midwest Supersite: Sensitivity Studies and Refinements. Department of Energy, Environmental and Chemical Engineering. Saint Louis, Washington University. **Master of Science:** 102.

Hildemann, L. M., G. R. Markowski, et al. (1991). "Chemical-Composition of Emissions from Urban Sources of Fine Organic Aerosol." Environmental Science & Technology **25**(4): 744-759.

Hogrefe, C., P. S. Porter, et al. (2006). "Temporal features in observed and simulated meteorology and air quality over the Eastern United States." Atmospheric Environment **40**(26): 5041-5055.

Jaeckels, J. M., M. S. Bae, et al. (2007). "Positive matrix factorization (PMF) analysis of molecular marker measurements to quantify the sources of organic aerosols." Environmental Science & Technology **41**(16): 5763-5769.

Lee, D., S. Balachandran, et al. (2009). "Ensemble-Trained PM2.5Source Apportionment Approach for Health Studies." Environmental Science & Technology **43**(18): 7023-7031.

Lee, J. H., P. K. Hopke, et al. (2006). "Source identification of airborne PM2.5 at the St. Louis-Midwest Supersite." Journal of Geophysical Research-Atmospheres **111**(D10).

Lough, G. C., J. J. Schauer, et al. (2005). "Summer and winter nonmethane hydrocarbon emissions from on-road motor vehicles in the Midwestern United States." Journal of the Air & Waste Management Association **55**(5): 629-646.

Marmur, A., J. A. Mulholland, et al. (2007). "Optimized variable source-profile approach for source apportionment." Atmospheric Environment **41**(3): 493-505.

Marmur, A., S. K. Park, et al. (2006). "Source apportionment of PM2.5 in the southeastern United States using receptor and emissions-based models: Conceptual differences and implications for time-series health studies." Atmospheric Environment **40**(14): 2533-2551.

Marmur, A., A. Unal, et al. (2005). "Optimization-based source apportionment of PM2.5 incorporating gas-to-particle ratios." Environmental Science & Technology **39**(9): 3245-3254.

Reff, A., S. Eberly, et al. (2007). "Receptor Modeling of Ambient Particulate Matter Data Using Positive Matrix Factorization: A Review of Existing Methods." Journal of the Air & Waste Management Association **57**: 146-154.

Rogge, W. F., L. M. Hildemann, et al. (1993). "Sources of Fine Organic Aerosol. 3. Road Dust, Tire Debris, and Organometallic Brake Lining Dust: Roads as Sources and Sinks." Environmental Science & Technology **27**: 1892-1904.

Rogge, W. F., L. M. Hildemann, et al. (1993). "Sources of Fine Organic Aerosol .4. Particulate Abrasion Products from Leaf Surfaces of Urban Plants." Environmental Science & Technology **27**(13): 2700-2711.

Rogge, W. F., L. M. Hildemann, et al. (1993). "Sources of Fine Organic Aerosol .5. Natural-Gas Home Appliances." Environmental Science & Technology **27**(13): 2736-2744.

Schauer, J. J., M. P. Fraser, et al. (2002). "Source Reconciliation of Atmospheric Gas-Phase and Particle-Phase Pollutants during a Severe Photochemical Smog Episode." Environmental Science & Technology **36**(17): 3806-3814.

Schauer, J. J., M. J. Kleeman, et al. (1999). "Measurement of emissions from air pollution sources. 1. C-1 through C-29 organic compounds from meat charbroiling." Environmental Science & Technology **33**(10): 1566-1577.

Schauer, J. J., M. J. Kleeman, et al. (1999). "Measurement of emissions from air pollution sources. 2. C-1 through C-30 organic compounds from medium duty diesel trucks." Environmental Science & Technology **33**(10): 1578-1587.

Schauer, J. J., M. J. Kleeman, et al. (2001). "Measurement of Emissions from Air Pollution Sources. 3. C1-C29 Organic Compounds from Fireplace Combustion of Wood." Environmental Science & Technology **35**(9): 1716-1728.

Sheesley, R. J., J. J. Schauer, et al. (2007). "Daily variation in particle-phase source tracers in an urban atmosphere." Aerosol Science and Technology **41**(11): 981-993.

Stolzel, M., F. Laden, et al. (2005). "Source apportionment of fine and coarse particulate matter and daily mortality in two US cities - A comparison of different methods." Epidemiology **16**(5): S95-S95.

Turner, J. R. (2007). St. Louis - Midwest Fine Particulate Matter Supersite Quality Assurance Final Report. St. Louis, Washington University in St. Louis**:** 34.

Turner, J. R. (2011). Speciated PM2.5 Data for the Midwest St. Louis Supersite. J. R. Turner. St. Louis, Sarnat, Stefanie Ebelt.

USEPA (2008). EPA Positive Matrix Factorization (PMF) 3.0 Fundamentals & User Guide. USEPA. Washington, DC, Office of Research and Development.

USEPA. (2010, March 5, 2010). "2002 National Emissions Inventory Data & Documentation." Clearinghouse for Inventories & Emissions Factors, Emissions Inventory Information Retrieved June 8, 2011, 2011.

USEPA. (2010, April 8, 2010). "Description of the EPA Unmix Model." Human Exposure and Atmopsheric Science (HEASD) Retrieved March 26, 2012, 2012, from http://www.epa.gov/heasd/products/unmix/unmix.html.

USEPA. (2010, April 24, 2012). "MOVES (Motor Vehicle Emission Simulator)." Retrieved May 8, 2012, 2012, from http://www.epa.gov/otaq/models/moves/index.htm.

USEPA. (2010, 5/13/2010). "Receptor Modeling." Technology Transfer Network Support Center for Regulatory Atmopsheric Modeling Retrieved 1/18/2012, 2012, from http://www.epa.gov/ttn/scram/receptorindex.htm.

USEPA. (2011, December 12, 2011). "SPECIATE Version 4.3." Clearinghouse for Inventories & Emission Factors, Software and Tools Retrieved March 7, 2012, 2012, from http://www.epa.gov/ttn/chief/software/speciate/index.html.

Watson, J. G., J. A. Cooper, et al. (1984). "The effective variance weighting for least squares calculations applied to the mass balance receptor model." Atmospheric Environment **18**: 1347-1355.

Zheng, M., G. R. Cass, et al. (2007). "Source apportionment of daily fine particulate matter at Jefferson street, Atlanta, GA, during summer and winter." Journal of the Air & Waste Management Association **57**(2): 228-242.